# Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies


RIGOROUS THESIS

# Google Econometrics: An Application to the Czech Republic


Author: **Mgr. Lukáš Platil**

Supervisor: **doc. Roman Horváth, Ph.D.**


Academic Year: **2014/2015**

## Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

The author grants permission to Charles University to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, September 11, 2015

Signature

## Acknowledgments

I am very grateful to the supervisor of my master thesis, doc. Roman Horváth, Ph.D., for his valuable comments, helpful suggestions, and advice throughout the process of writing the master thesis.

I would also like to thank to the referee of my master thesis, RNDr. Michal Červinka, Ph.D., for his comments, suggestions and questions for the master thesis defense, which I attempted to incorporate into the rigorous thesis.

# Abstract

This thesis examines the applicability of Google Econometrics – the use of search volume data of particular queries as explanatory variables in time series modeling – in the case of the Czech Republic. We analyze the contribution of Google data by comparing out-of-sample nowcasting performance and in-sample fit with control variables in three related areas: using an autoregressive model for unemployment, vector autoregression and logit models for GDP and household consumption, and Granger causality test for consumer confidence.

The improvement in quality of unemployment nowcasting is modest but statistically significant; sentiment index based on Google queries shows reciprocal relationship with the official Consumer Confidence Indicator, and it also provides superior nowcasts for household consumption as well as in-sample fit in logit models; its performance in GDP nowcasting is average among control variables. These conclusions proved stable also on an extended dataset.

In overall, the results suggest that Google Econometrics is applicable also to the Czech Republic, despite the fact that the internet penetration rate and Google popularity was lower over the analyzed period compared with developed economies where these methods were usually tested. In the future, Google data may be used together with other leading and coincident indicators to assess the current state of Czech macroeconomic variables that are available with a delay or with a lower frequency.



| | |
|---|---|
| **JEL Classification** | E21, E23, E24, E37 |
| **Keywords** | Google, search engine, search volume, search query, unemployment, sentiment, consumption |



| | |
|---|---|
| **Author's e-mail** | lukas.platil@gmail.com |
| **Supervisor's e-mail** | roman.horvath@gmail.com |

## Abstrakt

Tato diplomová práce se zabývá aplikovatelností 'Google ekonometrie' – neboli využitím dat o objemech vyhledávání určitých hesel jako vysvětlujících proměnných při modelování časových řad – v případě České republiky. Přínos Google dat analyzujeme za pomoci srovnání jak přesnosti 'out-of-sample' předpovědí, tak 'in-sample' kvality modelů oproti kontrolním proměnným ve třech oblastech: s využitím autoregresního modelu pro nezaměstnanost, vektorové autoregrese a logit modelů pro HDP a spotřebu domácností, a 'Granger causality test' pro důvěru spotřebitelů.

Zlepšení předpovědí v případě nezaměstnanosti je mírné, ale statisticky signifikantní; index důvěry založený na Google datech prokazuje vzájemnou provázanost s oficiálním indikátorem sentimentu, a zároveň přináší kvalitnější předpovědi pro spotřebu domácností i lepší 'in-sample' kvalitu logit modelů ve srovnání s kontrolními proměnnými; jeho přínos při modelování HDP je jen průměrný. Zjištěné závěry byly potvrzeny i po prodloužení časových řad.

Celkově výsledky naznačují, že 'Google ekonometrii' je možné aplikovat i na Českou republiku, a to i přesto, že po většinu analyzovaného období byla míra penetrace internetu i popularita vyhledávače Google menší ve srovnání s rozvinutými ekonomikami, na kterých byly tyto metody obvykle testovány. V budoucnosti tak mohou být Google data využívána spolu s dalšími vedoucími indikátory k odhadování současného stavu českých makroekonomických proměnných, které bývají zveřejňovány se zpožděním či nižší frekvencí.



| | |
|---|---|
| **Klasifikace** | E21, E23, E24, E37 |
| **Klíčová slova** | Google, vyhledávač, objem vyhledávání, heslo, nezaměstnanost, sentiment, spotřeba |



| | |
|---|---|
| **E-mail autora** | lukas.platil@gmail.com |
| **E-mail vedoucího práce** | roman.horvath@gmail.com |

# Contents

# List of Tables

# List of Figures

# Acronyms

| | |
|---|---|
| **ACF** | Autocorrelation function |
| **ADF** | Augmented Dickey-Fuller test |
| **AIC** | Akaike Information Criterion |
| **AR** | Autoregressive model |
| **ARX** | AR with additional explanatory variable(s) |
| **ARMA** | Autoregressive Moving Average |
| **ARMAX** | ARMA with additional explanatory variable(s) |
| **ARIMA** | Autoregressive Integrated Moving Average |
| **BCI** | Business Confidence Indicator |
| **BIC** | Schwartz Bayesian Information Criterion |
| **C-SI** | Case-Shiller Index |
| **CI** | Composite Confidence Indicator |
| **CCI** | Consumer Confidence Indicator |
| **CLIs** | Composite Leading Indicators |
| **CNB** | Czech National Bank |
| **CPI** | Consumer Price Index |
| **CZK** | Czech koruna |
| **CZSO** | Czech Statistical Office |
| **DJIA** | Dow Jones Industrial Average |
| **EU** | European Union |
| **GIIPS** | Greece, Italy, Ireland, Portugal, Spain |
| **GDP** | Gross Domestic Product |
| **HAC** | Heteroskedasticity and autocorrelation consistent |
| **HPI** | House Price Index |
| **HQIC** | Hannan-Quinn Information Criterion |
| **IPO** | Initial Public Offering |
| **KPSS** | Kwiatkowski-Phillips-Schmidt-Shin test |
| **MA** | Moving-average model |
| **MAE** | Mean Absolute Error |
| **MCSI** | Michigan Consumer Sentiment Index |
| **MSE** | Mean Squared Error |
| **OECD** | Organisation for Economic Co-operation and Development |
| **OLS** | Ordinary Least Squares |
| **PACF** | Partial autocorrelation function |
| **PRIBOR** | Prague InterBank Offered Rate |
| **PX** | Prague Stock Exchange Index |
| **RMSE** | Root Mean Squared Error |
| **SEM** | Simultaneous Equations Model |
| **SWIFT** | Society for Worldwide Interbank Financial Telecommunication |
| **U.K.** | United Kingdom |
| **U.S. / USA** | United States of America |
| **VAR** | Vector Autoregression |
| **VIX** | Chicago Board Options Exchange Volatility Index |

# TEZE RIGORÓZNÍ PRÁCE

(tvoří přílohu přihlášky ke státní rigorózní zkoušce)

**VYPLŇUJE UCHAZEČ:**

**Předpokládaný název rigorózní práce v češtině:**

| |
|---|
| **GOOGLE EKONOMETRIE: APLIKACE NA ČESKOU REPUBLIKU** |

**Předpokládaný název rigorózní práce v angličtině:**

| |
|---|
| **GOOGLE ECONOMETRICS: AN APPLICATION TO THE CZECH REPUBLIC** |

| | |
|---|---|
| **Předpokládaný termín předložení práce:** | 11. 9. 2015 |

**Charakteristika tématu a jeho dosavadní zpracování žadatelem** (rozsah do 1000 znaků)**:**

Many important macroeconomic variables are available with a significant delay or an insufficient frequency, which constitutes an inconvenient hurdle in assessing current economic conditions. Any variable that can be used to improve forecasting (or rather 'nowcasting') accuracy is of a particular interest. Leading or coincident indicators based on survey data are often used for such purposes, but they have some caveats (costly and time consuming collection, reliability).

Over the past decade, the development of new technologies has produced new types of electronic data – Google search volumes among others – that can reveal valuable information about economic agents. Such data have many advantages – freely and readily accessible, high frequency, can be analyzed almost in real time. This thesis examines the applicability of Google Econometrics – the use of search volume data of particular queries as explanatory variables in modeling economic time series – in the case of the Czech Republic.

This thesis has already been elaborated and defended as a diploma thesis with an excellent grade. Comments of the opponent and other improvements will be incorporated.

**Předpokládaný cíl rigorózní práce, původní přínos autora ke zpracování tématu, případně formulace problému, výzkumné otázky nebo hypotézy** (rozsah do 1200 znaků)**:**

So far, these methods have been tested mostly for developed Western economies (United States, Germany, France, etc.), but the application to developing countries or emerging markets is relatively scarce. These countries differ not only in the structure of their economy, but also in the internet penetration rate, which translates to the frequency of use and internet skills among population.

We analyze the central hypothesis – that Google data can be used also for modeling Czech macroeconomic variables, which has not been done up to date – in the areas of unemployment, consumer confidence, and overall economic situation, testing the following three sub-hypotheses:

a. Google search query data can be used to estimate the current Czech unemployment rate in advance compared with other methods.
b. Search query data can be used to assess the sentiment of consumers in the Czech Republic in advance compared with other (e.g. survey-based) methods.
c. An analysis of search query data can help predict economic development and crises in the Czech Republic more effectively compared with macroeconomic and financial data.

**Předpokládaná struktura práce** (rozdělení do jednotlivých kapitol a podkapitol se stručnou charakteristikou jejich obsahu)**:**

1. Introduction
2. Literature review

   *The second chapter will provide a review of the relevant literature on the topic, including the use of Google data in non-economic applications, the beginning and development in economic applications, and will also describe some of the potential disadvantages of such data.*
3. Data
   - 3.1 Google data
   - 3.2 Dependent and control variables

   *The third chapter will thoroughly describe the characteristics of available search query data, how to download them, aggregate them, and prepare for use in an econometric analysis. Moreover, the description of all other variables used in the thesis will be provided.*
4. Methodology
   - 4.1 Time series models
   - 4.2 Forecasting
   - 4.3 Logit model

   *The methodological chapter will describe all procedures and techniques used in the empirical section of the thesis. The description will provide overview of time series models including the Box-Jenkins methodology or the treatment of seasonality. We will provide a description of the creation of forecasts together with their evaluation and testing hypotheses regarding their predictive accuracy.*
5. Unemployment
6. Consumer confidence
7. Macroeconomic development

   *In the following three chapters, we will carefully build our empirical research on the foundations laid in previous chapters. In each of the analyzed areas, we will apply the models described in the methodological section, evaluate them and provide a discussion together with a comparison with the reviewed literature.*
8. Conclusion

**Vymezení podkladového materiálu** (např. analyzované tituly a období, za které budou analyzovány) **a metody (techniky) jeho zpracování:**

On the Google Trends website (www.google.com/trends), Google provides information about volumes of searches of individual queries conducted through its search engine. The information provides a measure of the popularity of individual search queries relative to the total number of searches conducted in a given geographical location at the time. All time series start from January 2004, and we will analyze the data sample for years 2004-2014.

The contribution of Google data will be tested by a comparison of out-of-sample nowcasting performance (measured by the mean square error), as well as in-sample fit, with the performance of control variables. A suitable time series model, found e.g. through the Box-Jenkins methodology, will be used for each of the analyzed dependent variables; the models will include autoregressive processes (AR), autoregressive integrated moving averages (ARIMA), vector autoregressions (VAR), or logit models for dichotomous variables.

Further, such models will be augmented with Google data and Clark-West (CW) test and modified Diebold Mariano (MDM) test of equal predictive accuracy for nested and non-nested models will be used to assess the contribution of search query data compared with benchmark models and with models of control variables. In addition, Granger causality test can assess an interconnection of several tested variables.

**Základní literatura** (nejméně 10 nejdůležitějších titulů k tématu a metodě jeho zpracování; u všech titulů je nutné uvést stručnou anotaci na 2-5 řádků)**:**

□ Bańbura, M., Giannone, D., Reichlin, L. (2010): Nowcasting, European Central Bank, Working Paper No. 1275, December 2010

*The authors defined the term "nowcasting" as "a forecast of the present, very near future and very recent past", an activity "relevant especially for macro variables that are collected with a low frequency (typically quarterly) and that are published with a substantial lag." They designed a model producing nowcasts in relation to real time releases of various economic data, and also studied the interrelation between the consumer confidence and the GDP growth in the Euro Area.*

□ Choi, H., Varian, H. (2012): Predicting the Present with Google Trends, The Economic Society of Australia, *The Economic Record*, Vol. 88, Special Issue, pp. 2-9, June 2012

□ Choi, H., Varian, H. (2009): Predicting the Present with Google Trends, Google Inc., April 2009

*Choi and Varian were the first authors to use Google data in economics. The aim of their article was to introduce and familiarize readers with Google Trends, and to show examples of the use of such data. More specifically, they analyzed automotive sales, retail sales and housing market in the United States. Their invite found a significant response and a whole field of "Google Econometrics" has emerged. They later extended their study in the article from 2012.*

□ Clark, T.E., McCracken, M.W. (2011): Advances in Forecast Evaluation, Federal Reserve Bank of Cleveland, Working Paper 11-20, September 2011

*This paper is a survey on recent developments in the evaluation of point forecasts. The authors described forecasts evaluation methods, including advancements in the evaluation of forecasts at the population level, in the finite sample, and the evaluation of conditional versus unconditional forecasts. Moreover, the authors presented some original result, e.g. in the area of optimization of power in determining the split of a sample into in-sample and out-of-sample portions.*

□ Clark, T.E., West, K.D. (2007): Approximately Normal Tests for Equal Predictive Accuracy in Nested Models, *Journal of Econometrics*, Elsevier, vol. 138(1), pp. 291-311, May 2007

*The authors proposed a test of equal predictive accuracy for nested models – a situation when a parsimonious model is compared to a larger model that nests the first one. Under the null hypothesis the parsimonious model generates the data, so the larger model introduces a noise into its forecast. The authors proposed a method how to adjust the mean square error of the larger model and then test whether the adjusted mean square error difference is zero.*

□ Da, Z., Engelberg, J., Gao, P. (2013): The Sum of All FEARS: Investor Sentiment and Asset Prices, Working paper for The Journal of Finance, 2013

*The authors created their own index of investor sentiment based on Google data. They started with a large set of words that had economic meaning and, at the same time, had either positive or negative connotation. After conducting analytical procedures on the data, they ended up with an index calculated as an average of search volumes of 30 queries. Since all constituent parts of the final index had a negative correlation with market returns, they called it FEARS.*

□ Della Penna, N., Huang, H. (2009): Constructing Consumer Sentiment Index for U.S. Using Internet Search Patterns, University of Alberta, Department of Economics, Working Paper No. 2009-26, 2009

*The authors also created their own index of consumer sentiment based on Google data. For their analysis, they chose categories of queries expected to be related to individual questions asked in the official survey for the University of Michigan Consumer Sentiment Index. After the selection procedure, they made an index as an average of volumes of search of four categories: bankruptcy, luxury goods, energy & utilities, and office furniture.*

□ Diebold, F.X., Mariano, R.S. (1995): Comparing Predictive Accuracy, *Journal of Business & Economic Statistics*, American Statistical Association, vol.13 (3), pp. 253-64, July 1995

*Diebold and Mariano proposed a relatively simple test for comparing the quality of two competing forecasts, allowing the use of a wide variety of accuracy measures. The comparison is model-free, meaning that the test statistic takes into account only forecasts (or forecasting errors), disregarding the models that generated them (they may not even be known). For this reason, this test is often used for a comparison of quality of forecasts of non-nested models.*

□ Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L. (2009): Detecting Influenza Epidemics Using Search Engine Query Data, *Nature*, Vol. 457, 19 February 2009

*A seminal paper in the area of Google data application – the first use of Google data in making short-term predictions of real-life variables – was conducted in the area of medicine. The authors examined search queries related to influenza-like illnesses and their symptoms, and found a tight correlation between the development of the volume of searches of such terms and a subsequent number of visits of physicians by people having influenza-like symptoms.*

□ Hansen, P.R., Timmermann, A. (2012): Choice of Sample Split in Out-of-Sample Forecast Evaluation, CREATES Research Papers 2012-43, February 2012

*The authors demonstrate that out-of-sample forecast evaluation results may critically depend on how the split between in-sample and out-of-sample portions is determined. They propose a test statistic to deal with size distortions, and recommend to make the split relatively early in the sample. Also, they make arguments for and against available forecasting schemes: fixed, rolling, and recursive window schemes for parameters estimation.*

□ Harvey, D.I., Leybourne, S.J., Newbold, P. (1997): Testing the Equality of Prediction Mean Squared Errors, Elsevier, *International Journal of Forecasting*, vol. 13, pp. 281-291, June 1997

*The authors analyzed the behavior of two possible tests of a hypothesis of equal predictive accuracy together with their modifications that were designed to circumvent shortcomings in the original formulations. As a result, the authors proposed several changes to the original Diebold Mariano test, such as an adjustment of the test statistic for the length of the forecast horizon and the number of forecasts, and demonstrated better size properties of the modified version.*

# Rigorous Thesis Introduction

The presented rigorous thesis is an extension of my diploma thesis that was successfully defended with a grade A in June 2014 at the Institute of Economic Studies (Faculty of Social Sciences, Charles University in Prague). The original thesis has been revised and some parts have been modified to reflect the comments and suggestions of the referee. More importantly, the text has been extended by several sections to answer inquiring questions of the referee. The principal upgrade of this rigorous thesis is an extension of the dataset to analyze the stability of the original results.

The thesis is concerned with the topic of Google Econometrics, which is the use of standard techniques of time series modeling while using volumes of Google searches of particular terms as additional explanatory variables. In the original thesis, this was applied to the topics of unemployment, consumer sentiment, and overall economic situation, on a dataset from January 2004 to December 2013 in the Czech Republic, an analysis which had not been done up to the publication of the thesis.

In this rigorous thesis, we update the original results by extending the dataset to December 2014 (that is by 12 monthly observations or 4 quarterly ones) for all tested variables. A complete analytical procedure of this dataset has been conducted like in the original case. Appendix C contains the most important results of this analysis, and Chapter 8 describes and compares new results with the original ones and discusses their implication for a potential use in the future.

The literature review in Chapter 2 has also been updated to reflect the application of Google Econometrics to economies in several new countries. Most importantly, during the last year, two applications of Google Econometrics to the Czech Republic have appeared; this is discussed in section 2.2.7. Most notably, the literature review has been extended by section 2.4 which discusses the use of other internet-users based data in economics – social media feeds represented by Twitter; this was one of the questions proposed by the referee.

The second question of the referee, which aimed at the practical aspects of the application of Google Econometric in real life due to variations in Google data depending on the day of download, is briefly discussed in section (8.4). Lastly, while we do recognize the extensive length of the original thesis as well as of this rigorous thesis for any potential reader, we didn't see the point of shortening it as we see all three focal points of the analysis to be relevant and contributive. Rather, as the sections 5, 6 and 7 can be read relatively independently, we would recommend the reader to focus on the topic he finds the most intriguing.

# 1 Introduction

Many important macroeconomic variables are available with a significant delay. For example, quarterly values of Gross Domestic Product in the Czech Republic are published with a delay of almost three months, and they are also revised later on. These variables are used to assess the state of the economy, and since evaluating current economic conditions is an important part of decision making about macroeconomic policies (by governments, central banks), the delay in publishing such data constitutes an inconvenient hurdle.

Publication frequency may also seem insufficient, that is one of the reasons why the importance of early available data rises during times of macroeconomic turmoils or isolated shocks. For example concerning the recent economic crisis which started with a burst of a bubble in the American housing market, the current situation is closely followed as it may have implications for the state of the whole economy (Wu & Brynjolfsson, 2013) – but official housing market indicators in the U.S. are available with a delay of 2 or 3 months.

Having reliable estimate of current values is useful. Bańbura et al. (2010) define the word 'nowcasting': "*it is a forecast of the present, very near future and very recent past; this activity is relevant especially for macro variables that are collected with a low frequency (typically quarterly) and that are published with a substantial lag.*" Making forecasts is one of the basic motives for creating economic models, and any variable that can be used to improve forecasting accuracy is of a particular interest.

So called leading or coincident indicators are usually used for the purpose of nowcasting. These indicators are often based either on economic variables that are published earlier and/or with a higher frequency than the nowcasted variable, or on data coming from surveys among economic agents. Such surveys usually investigate the economic situation of respondents or their assessment of the current and future overall economic situation. But surveys have some caveats: they study only limited sample of economic agents; they are costly; it takes some time to carry them out. Moreover, usually no incentive exists for respondents to provide truthful answers.

Together with technologic development during past decades, new type of electronic data has appeared that can also reveal valuable information about human behavior (Wu & Brynjolfsson, 2013). These data, capturing the behavior of many economic agents for example on the internet, can be analyzed almost in real-time and therefore have big potential when assessing current economic conditions. The internet has become an integral part of almost every aspect of human life in developed countries (Hohenstatt et al., 2011), many people use it to search for information of any kind.

Google is the most popular search engine around the world – so popular that the word 'google' itself in English language has become a synonym for conducting search on the internet using search engine. In 2009, Google started to provide information about volumes of searches of individual queries; the data are available on the website Google Trends (www.google.com/trends, earlier called Google Insights for Search). The information provides a measure of the popularity of individual search queries relative to the total number of searches conducted in a given geographical location at the time; time series start in January 2004.

Given its popularity around the world (dominating position in most of the developed countries), Google provides a broad range of data that would have been costly to attain otherwise; it is also available almost in real-time and it is freely and easily accessible. Researchers from various scientific areas quickly started to analyze the data to inquire into the relationship between volumes of searches and particular real-life variables; one of the first uses was for detecting potential influenza epidemics in the United States based on the popularity of searches connected to flu symptoms.

Similar methods have quickly spread also to economics, forming so called Google Econometrics – that is the use of classic methods of time series modeling while using Google data as additional explanatory variables. It is based on the idea that by entering a particular term into search engine, internet users reveal information about their current interests; from an economic point of view, this may be connected to their economic situation, overall economic conditions or even direct consumption plans.

For example Schimdt & Vosen (2011) characterize that "*while macroeconomic variables can show consumers' "ability to spend" and data from surveys can show consumers' "willingness to spend", Google data can directly reveal consumers' "preparatory steps to spend"*." Because as Wu & Brynjolfsson (2013) add, every time a consumer searches for a product on the internet, he also reveals potential intentions to make a transaction.

Consumption is not the only possible area of research; people may also reveal their economic situation – for example when searching for queries related to jobs, we can analyze the relation to the unemployment rate. Similarly when examining queries related to debt, we can assess their financial situation or general sentiment; also, particular queries reveal attention paid to some company and so on.

Indeed, almost every study conducted on the topic of Google Econometrics showed that Google data can improve forecast accuracy of benchmark models and given their early availability, it is suitable for nowcasting or short-term forecasting; this was analyzed for unemployment, consumption, housing market, financial market or even overall economic development. So far, most of the studies were done using data from developed economies, such as the United States, Germany, France and other Western European countries.

Application of Google Econometrics to developing countries or emerging markets is relatively scarce, but the researchers have found the contribution of Google data even there. For example for Israel and various sectors of the economy (Suhoy, 2009), for Chile and automotive industry (Carrière-Swallow & Labbé, 2010), for unemployment in Turkey (Chadwick & Sengul, 2012) and China (Su, 2014), or for India and real estate stocks (Das & Ziobrowski, 2015). Compared with developed economies, these countries differ in the internet penetration rate – this would translate both to the frequency of use of the internet as well as internet skills; the interconnection between behavior of people on the internet and real economy is not automatic.

In this thesis, we examine the applicability of Google Econometrics to the economic data of the Czech Republic, something which has not been done prior to the publication of the thesis,[1] over the period 2004–2013. Similarly to the aforementioned countries, the Czech Republic had lower internet penetration rate for most of the analyzed period compared for example with states of Western Europe; and more importantly, Google does not dominate the Czech search engine market, which also plays a role in the data representativeness.

We tested the central hypothesis – that Google data can be used also for modeling Czech macroeconomic variables – in the three following areas: unemployment; consumer confidence; and overall economic situation. For unemployment, we used a framework of ARIMA models to create out-of-sample nowcasts; then, the Clark-West test of equal forecast accuracy for nested models was used to assess the contribution of search query data compared with both the benchmark ARIMA model as well as with models augmented with control variables.

[1] Since the publication of the original master thesis, two applications of Google Econometrics to the Czech Republic have appeared, and are described in the literature review section.

To analyze consumers' confidence, we created own Google Consumer Sentiment Index based on search query data, combining two approaches used in the related literature (Della Penna & Huang, 2010, Da et al., 2013). We connected chosen queries to specific questions asked during surveys for official indicators, and compared our index with the official one to see if it is possible to assess consumer confidence in advance to surveys, using in-sample correlation coefficients and Granger causality test.

Finally, we tested the performance of our Google index in forecasting overall economic situation (GDP growth) and household consumption. This was firstly done comparing in-sample fit in forecasting economic downturns and below-average growth of household consumption in the framework of logit models, comparing Google data with the set of control variables. Secondly, we modeled the growth of GDP and household consumption with vector autoregression (VAR) and compared out-of-sample nowcasting accuracy of VAR models with different variables; this was done using the Clark-West test for nested models and Modified-Diebold-Mariano test for non-nested models.

The thesis is structured as follows: Chapter 2 provides brief literature review; Chapter 3 describes the data and Chapter 4 the methodology used; Chapter 5, 6, and 7 describe empirical research and results for the analysis of unemployment, consumer confidence, and overall economic situation. Chapter 8 comments on the updated results of the extended dataset. Chapter 9 concludes. In addition, Appendix A provides a brief statistical study on the internet use in the Czech Republic, as well as the popularity of Google compared with other countries.

# 2 Literature review

## 2.1 Google data outside of economics

The first use of Google data for making short term prediction of a real-life variable was conducted by Ginsberg et al. (2009) in the area of medicine. More specifically, they made use of the concern of internet users with their own health when searching for solutions of their medical issues. This seminal paper in the area of Google data application was a joint study of employees of Google Inc. and the Centers for Disease Control and Prevention of the United States. For the U.S. data for years 2004–2008, they examined search queries related to influenza-like illnesses and their symptoms; they found a tight correlation between the development of the volume of searches for such words and the subsequent number of visits of physicians by people having influenza-like symptoms in the next few days.

They concluded they were able to successfully nowcast current levels of influenza activity in the U.S. with a delay of one day, compared to official data published with a delay of two weeks. The results of this study were published in the journal Nature and Google also launched a webpage that analyzes current trends of flu activity around the world (www.google.org/flutrends). The use of Google data quickly spread also to other areas, including economics. Firstly, we provide few examples from other fields.

Yang et al. (2010) also studied medical issues – they used volumes of search for queries connected to mental health as a proxy for the occurrence of depressions. They examined the data for 2004–2009 for 54 geographical locations around the world and found that there is a seasonal occurrence of depression with a significant correlation with temperature oscillations; the pattern was clearer for locations farther from the equator.

Data about searches made on the internet were also analyzed in social sciences. Ripberger (2011) used these data in politology for the U.S. as a proxy for public attention paid to publically debated issues (topics such as health care, global warming, and terrorism). He showed that this proxy is valid in comparison to usual methods of public attention estimation, such as coverage in New York Times.

In sociology, the data was used for example by Stephens-Davidowitz (2013) for the U.S. and years 2004–2007. He analyzed searches for racially affected words to attain an indicator of racial animus in individual states of the USA and found that racial animus cost Obama approximately 4 percentage points of votes on the national level during 2008 and 2012 elections. Racial animus is the kind of information that is unlikely to be credibly obtained with survey-based methods; many people would rather prefer to give a socially acceptable answer.

## 2.2 Google data in economics

Since Google is the source of the data, it is not surprising that the authors of the first article in economics were also from this company: the Chief Economics., H. Varian, and the Senior Economist, H. Choi. They wrote their article (Choi & Varian, 2009; later updated in Choi & Varian, 2012) with the aim to introduce and familiarize readers with Google Trends and to show examples of the use of the data. They made short-term forecasts of economic indicators using the data to motivate other researchers to continue in the field.

From the brief description in the introduction chapter, it is clear that their intentions found response and the whole field of Google Econometrics emerged. In the paragraphs below, we describe the topics discussed by Choi & Varian (2009, 2012) and their approach; also, we shortly review articles by other authors that either extended the topic or applied the same methods to data of other countries. More detailed description of studies most relevant to the topics examined in our thesis is included in appropriate chapters. Generally, Choi & Varian (2009) put en emphasis on the interconnection between what people search for on the internet and sales in several sectors: automotive sales, retails sales and housing market.

Other authors also used electronic data to analyzed the current economic activity, for example Döhrn (2013) used electronic toll data for Germany to find a relation between the number of kilometers driven by vehicles above 12 tons of weight and industrial production (and indeed, models using such data had lower forecasting error compared with survey-based indicators). Gill et al. (2012) used the data about electronic transactions – the flow of wholesale payments between banks through SWIFT and data about payments with debit and credit cards.

This type of data reveals real activity of economic agents, but does not capture all activity – for example, flow of payments trough SWIFT does not take into account payments between clients of the same bank, data from debit and credit cards do not take into account cash payments. On the other hand, as Vosen & Schmidt (2011) characterized, "*Google data can directly* *reveal consumers' preparatory steps to spend*". The internet is used to search for information especially about purchases that are more financial demanding, for example Wu & Brynjolfsson (2013) provide a statistics that during 2012, 90% of home buyers in the United States used the internet during the process of purchase.

### 2.2.1 Housing market

Choi & Varian (2009) tested this on the statistics about New Houses Sold and For Sale for the United States over the period 2004–2008. Typically, the approach of these authors was modeling the initial indicator with a baseline seasonal autoregressive process, augmenting it with Google data, and comparing the quality of out-of-sample one-step-ahead forecasts using Mean Absolute Error (MAE). For housing market, they used categories of Google queries "Real Estate" and its sub-category "Real Estate Agencies". With the process described above, they found that Google data improved the MAE of the benchmark by 12%.

The same topic, also for the United States, was later examined by Kulkarni et al. (2009), Hohenstatt et al. (2011) and Wu & Brynjolfsson (2013). They compared the contribution of Google data (either categories or individual search queries) with indicators usually used to assess the situation in the U.S. housing market, such as Case-Shiller Index (C-SI) or House Price Index (HPI). Kulkarni et al. (2009) studied this market for 20 U.S. cities for years 2005–2009. Using Granger causality test, they found that relevant Google data 'Granger caused' both C-SI and HPI, but not vice-versa. Hohenstatt et al. (2011) also showed for the data for years 2004–2009 that Google index predicted the volume of transactions and housing prices (C-SI), but there was also a reverse causality creating an endogeneity problem.

Wu & Brynjolfsson (2013) modeled the number of transactions with the data for 2006–2011. They used an autoregressive model with additional explanatory variables (including HPI). Using Google data improved the error of predictions by 2.3% (nowcasting) and 7.1% (forecasting), and when compared with experts' forecasts (from the National Association of Realtors), the improvement was 23.6%; Google data could also predict HPI itself.

Because using Google data had bigger gains for forecasts than for nowcasts, this study partially confirmed expectations of Hohenstatt et al. (2011) that the advantage of Google data in this area is a longer period elapsed (up to 12 weeks) between the time the search for information begins and the time people make the decision about potential purchase of a house; so it can be used to predict the situation in this market several months ahead.

### 2.2.2 Car sales

Purchase of a car is usually also a large investment which requires more attention. Choi & Varian (2009) studied the sales using the statistics 'Motor vehicles and parts dealers' and also 'US car and light-truck sales by make'. To model the sales of individual brands, they used the data from subcategories "Automotive / Vehicle Brands" and presented results for Ford, Chevrolet and Toyota. They did not find any contribution of Google data when forecasting Chevrolet sales, but the improvement was 3% for Ford and 12% for Toyota.

Choi & Varian (2012) later updated their results, analyzing a longer period and modeling 'Motor Vehicles and Parts Dealers' sales' as a whole, using categories of queries "Truck & SUVs", "Automotive Insurance" and "Motorcycles". They found that while such data improved forecast accuracy by 10.5% during the whole period, it was 21.5% when only the recession period (December 2007 to June 2009) was taken into account. They concluded that Google data capture well the dynamics during times of unexpected changes that simple autoregression cannot predict.

Similar observation in the same area, but for an emerging market, was made by Carrière-Swallow & Labbé (2010). They also found that Google data – searches for most popular car brands in Chile – improved the forecast accuracy of automotive sales compared with models with control variables (index of economic activity); and the data also captured well turning points in car sales during the period 2006–2010. This study was one of the first applications of Google Econometrics on emerging markets, showing that it can be used even there – at least for sales of goods that represent significant spending expenses. As the authors note, it is not automatic that internet would be a common part of consumers' behavior in these countries with lower internet penetration rate.

### 2.2.3 Retail sales

Other authors modeled also other categories of retails sales or retails sales as a whole, trying to improve the performance of survey-based indices. Kholodilin et al. (2010) and Schimdt & Vosen (2011) for the United States, Chamberlin (2010) for the United Kingdom or Schmidt & Vosen (2012) for

Germany. They typically used categories of search queries related to consumption behavior (either theoretically or empirically), aggregated the data (for example by principal component analysis) and compared the in-sample fit or the forecasting performance with models containing control variables, such as survey based indicators or financial and macroeconomic data.

The authors usually found that the quality of models with Google data was significantly superior. For example Kholodilin et al. (2010) used categories including "motor vehicles and parts", "transportation services", "financial services and insurance", or "health care" and improved the forecasting accuracy (Root Mean Squared Error, RMSE) by 17% compared with a baseline model, 8 p.p. more than when using control variables. Schmidt & Vosen (2011) arrived at even bigger improvement (RMSE), 64% compared with autoregressive process and 21% compared with models with control macroeconomic variables. In addition, the authors found that only Google data captured the turning point in retails sales in December 2008.

Chamberlin (2010) analyzed individual categories of retails sales in the United Kingdom in a period 2004–2010. He always matched a retail trade category to an appropriate category of search queries and discussed the statistical significance of coefficients in estimated autoregressive models for first differences of the data. His results were contradictory – he found significant coefficients for example for "food and shopping", "department stores", or "clothes and footwear"; but not for example for "electronic household appliances" or "audiovisual devices and records".

And Schmidt & Vosen (2011) used categories of "travel and transportation", "education", "alcoholic beverages and tobacco products", "telecommunication", "food and non-alcoholic beverages", or "restaurant services" to create their own index of private consumption for Germany. Model containing Google data improved forecasts (MSE) by 3 to 71% compared with competing models; control variables included consumer confidence indices or macroeconomic data (real income, interest rate).

### 2.2.4 Consumer confidence

Consumer confidence was also studied by Choi & Varian (2012) – the Roy Morgan Consumer Confidence Index for Australia. Using Bayesian method of spike and slab regression to determine which categories of Google data explain this index, they found that "Crime & Justice", "Trucks & SUVs", and "Hybrid & Alternative Vehicles" improved the MAE of an autoregressive process by 9.3%. Analysis of sentiment of economic agents was conducted also by other authors, for example by Della Penna & Huang (2009) who used Google data to predict consumption in the United States; or Becchetti et al. (2012) who studied the interconnection of sentiment and the threat of financial crisis. A more detailed description of their methods and results is provided in Chapter 6 concerned with this topic.

Da et al. (2013) and Beer et al. (2013) also created their own sentiment indicator based on Google data and used it to explain and predict movements in the financial market (in the United States and France, respectively). This is also one of the areas of research where Google Econometrics has been quite popular. Usually, information about search volumes is used to measure two phenomena – investors' sentiment (like in the articles mention above) or the level of attention investors pay to individual stocks.

### 2.2.5 Financial markets

For similar purposes, market-based indicators are often used – such as the volume of trade, first day IPO returns, volume of IPOs, implied volatility, or mutual funds flow. Even though such data offer early availability and high frequency of observations, it is often criticized as being a result of more market forces, not sentiment or attention of investors only (Da et al., 2013).

Da et al. (2011), Vlastakis & Merkellos (2012) and Preis et al. (2010) used the volume of search either of names of individual companies or tickers of stocks traded at the U.S. stock exchanges; they used abnormal search volume as a proxy for unusual attention of investors. They discovered a connection between the attention and the volume of trade and volatility of returns, while the impact on price growth was not clear. Takeda & Wakao (2014) found the same for Japanese stocks – the correlation between search intensity and trading volume was strongly positive over the period 2008–2011.

Ding & Hou (2011) arrived at a similar conclusion based on an analysis of the U.S., U.K., European and Chinese data. In addition, they discovered a connection between the attention of retail investors, the width of investors' base and stocks' liquidity. Bank et al. (2010) observed the same relation to liquidity also for German data, and similarly to Da et al. (2011), they found that increased attention was followed by a short-term rise in prices that was followed by a reversal to initial values.

Krištoufek (2013) used the observation about attention and return volatility and designed a portfolio management strategy where the weights of individual stocks were inversely proportional to their popularity measured by the volume of search. Examining the U.S. data, he concluded that this strategy performed better than the benchmark Dow-Jones Index as well as uniformly weighted portfolio. Challet & Ayed (2014) criticized approached usually applied, claiming that changes in search volume could mean too many things to use it as a proxy for attention without any additional adjustment – it could be good news, bad news, or it may have nothing to do with financial markets.

Das & Ziobrowski (2015) used a more general approach to model realty stocks in India. They used Google queries related to real estate to forecast the returns of publicly traded real estate stocks. They found that when taking into account co-movements with the S&P Bombay Stock Exchange Sensitivity Index, relevant searches were significantly interconnected with future changes in stock returns and their conditional volatility.

### 2.2.6 Unemployment

Another area that Choi & Varian (2012) discussed was unemployment. In the United States, a statistics about Initial claims for unemployment benefits is closely followed as a leading indicator for unemployment rate, and their goal was to predict changes in this variable with search query data. They used two categories: "local / jobs" and "society / social services / welfare & unemployment" and found that Google data improved out-of-sample forecasts (MAE) of the benchmark autoregressive process by 6%; but again when looking only at the recession period from December 2007 to June 2009, the improvement was more than 13%.

Among other aspects of the economy, Suhoy (2009) analyzed also unemployment for Israel using a sub-category of Google data "Human resources". She found that this sub-category had the biggest explanatory power not only for unemployment, but also for other sectors of Israeli economy, such as industrial production. This was another example of a successful application of Google Econometrics in an emerging market.

Su (2014) provided relevant evidence for China, where official unemployment figures are published only quarterly and even these are considered unreliable and misleading (because of the official definition of unemployment and the data collection procedures). The author created two indices (one based on data from Google, the other on data from a more popular Baidu engine) and found that they catch well structural breaks in October and November 2008. They were also significantly correlated with Purchasing Managers' Employment Indices and other macroeconomic variables connected to real unemployment rate, and could also be used to improve their forecasts.

Also other authors analyzed unemployment: D'Amuri (2009, Italy), Askitas & Zimmermann (2009, Germany), D'Amuri & Marcucci (2010, United States), Chadwick & Sengul (2012, Turkey), Fondeur & Karamé (2013, France), or Tuhkuri (2014, Finland); more detailed description is provided in Chapter 5.

### 2.2.7 Czech Republic

Since the publication of the original master thesis, two newly published studies analyzed the possibility to use Google data also in the case of the Czech Republic. On a data sample from January 2007 to October 2014, Saxa (2014) analyzed mortgage lendings; more specifically the nominal volume of new mortgages provided to households in the Czech Republic. He found a strong correlation between this series and Google searches of mortgage related terms with a lag of two (0.75) and three (0.74) months.[2]

When modeling mortgage lendings by an autoregressive process and incorporating Google data, he found an improvement in both in-sample fit and out-of-sample forecast accuracy. The in-sample fit, measured by adjusted $R^2$, increased from 0.05 to 0.39. The out-of-sample forecasting accuracy, measured by MAE and RMSE, improved by 18% and 23% respectively, for one month ahead forecasts. The results were slightly worse when accounting for seasonality in the data.

Moreover, based on his search index, the author created an indicator of banks' unwillingness to lend. He calculated it as a difference between the growth of search of terms related to mortgages (a proxy for credit demand) and the growth in mortgage lendings two months later. He found that this indicator captured well the evolution during the end of 2008, when the gap between the demand and supply of mortgages rose, as banks implemented restrictive measures.

Pavlíček & Krištoufek (2015) analyzed the potential to used job-related searches to forecast unemployment rate in Visegrad Group countries – the Czech Republic, Slovakia, Poland and Hungary – between January 2004 and December 2013. They found that for the in-sample fit, the incorporation of Google data strongly improved models for all countries (measured by adjusted $R^2$), with Czech Republic having the most promising results.

[2] The exact string used was "hypotéka + hypoteka + hypotéky + hypoteky + hypoteční + hypotecni + hypotéku + hypoteku + ("'úvěr na bydlení") + ("uver na bydlení")".

In an out-of-sample exercise, the results were mixed – while Google data outperformed the baseline model for the Czech Republic for all 3 lag specifications (3 lags, 6 lags, and 9 lags used as explanatory variables), this was true only for 2 specifications for Hungary, 1 for Poland and none for Slovakia. The authors identify labor market characteristics as s possible causes of these different results; for example the willingness or desire to move for work abroad among Poles or Slovaks compared with Czechs and Hungarians.

## 2.3 Disadvantages

It is clear that Google econometrics has been successfully applied to economies of various countries and to various topics. Of course, there are some caveats in the use of these data. One of them is a rather short time series of Google data – currently 10 years. This is a problem for some models that estimate coefficients over a longer time span or for backward testing of Google data appropriateness. On the other hand, this can only improve with time. Another question is the representativeness of the data. As Gill et al. (2012) note, different age (but also income, etc.) categories differ in the internet penetration rate as well as in ways they use the internet.

One of the most important things is the choice of Google query. Firstly, it has to be used comparably over the whole analyzed period – some words may change their meaning or be replaces (Askitas & Zimmermann, 2009). This can be further amplified by the shift of the focus on the internet use to entertainment and social media (Suhoy, 2009). As Saxa (2014) pointed out, this may even change during the course of the years. In his analysis of the Czech Republic, he found a special type of seasonality in Google data – a decrease in the relative volume of searches of some terms as people start to conduct more Christmas related searches.

Also, even if the word is correctly chosen, it may be used by people that are not in the center of attention. For example as D'Amuri (2009) notes, both unemployed and employed people can search for "jobs" on the internet. He claims that in Italy, the fractions of these two groups is 3:1, and their anticipated cycles are opposite, which can introduce noise into the data. Similarly, students or retirees may search for a job on the internet and would not count amongst unemployed. Next, if the use of Google data becomes more popular, there is also a threat of data corruption (using some mechanical searching method to introduce more noise in the data) (Wu & Brynjolfsson, 2013) – and it will be necessary to consider which sectors are prone to a potential abuse.

## 2.4 Twitter data

Google is not the only source of internet data that we can use in economics. Mao et al. (2011) distinguish three main sources of internet data that have been analyzed in economic models. Firstly, (1) news media – either on-line versions of traditional newspaper articles (such as Wall Street Journal columns), or specifically on-line sources. When using such data, the assumptions is that either the information contained forms the sentiment of readers or that it reflects the demand for information – a measure of attention to a particular phenomenon. For example Gerrow & Keane (2011) analyzed the content of internet articles about the situation in stock exchanges, taking news data as a proxy of investors' sentiment.

The second source is (2) web search data, most notably Google, which is thoroughly analyzed in the thesis. In that case, researchers aim to inquire about the current economic interests, needs, or plans of economic agents from the type of queries users process. The third source of data is (3) social media feeds, such as internet message boards (either general or topical), analyzed e.g. by Antweiler & Frank (2004), weblog entries analyzed e.g. by Gilbert & Karahalios (2010), number of modifications of Wikipedia pages of individual companies as a proxy for attention by Rubin & Rubin (2010), or lately the most popular platforms – Twitter and Facebook. Unlike the previous cases, where the data covers either passive consummation or search for information, this source of data captures an active expression of ideas of individual users in a complex form.

Twitter, a website launched in 2006, offers social networking services to users posting messages with a length of up to 140 characters. People can follow individual users or topics (usually denoted with a hashtag "#"), react, retweet, or interact with other users in other ways. Unless restricted by the user, posted messages ("tweets") are publicly available in real time, and researchers can access samples from archives of all tweets through Twitter content resellers. With about 300 million active users and 500 million tweets per day,[3] as Janetzko (2014) notes also about other social media, Twitter rather produces an overabundance of data. It is therefore a challenging task to find a way to extract valuable information from this source.

As Antenucci et al. (2014) and Culotta (2010) note, one of the possible advantages of Twitter data compared with Google is the public availability of tweets and the fact that it often contains some metadata about individual

[3] http://www.businessinsider.com/twitter-tweets-per-day-appears-to-have-stalled-2015-6

users (city, age, sex, etc.). While Google publishes its data already in a processed form as a relative search volume index that can only be restricted to a given geographical area, Twitter data allow researchers to do their own processing both in terms of tweets content as well as possible extraction of information about individual posters.

### 2.4.1 Contemporaneous events

One of the popular applications of Twitter data, both in academic and commercial sphere, has been the analysis of mood of Twitter users based on the linguistic analysis of tweets. For example, Golder & Macy (2011) mapped the evolution of mood during a day, finding that people wake up in a good mood, which gradually deteriorates during the day and improves later in the evening. This evolution is also dependent on the season (the length of daylight) and the day of the week, people being happier on weekends.

Similarly, researchers from the Northeastern University and the Harvard University created "The Pulse of Nation: U.S. Mood Throughout the Day inferred from Twitter" visualization of the mood of US Twitter users, based on an analysis of Tweets  from September 2006 to August 2009, mapping regularities in the sentiment of users in different U.S. states across the continent.[4] Some other websites attempt to monitor similar characteristics, and also some commercial applications of Twitter mood analysis have been launched, such as a "Twitter Mood Light", which is a box than changes its color based on the predominant sentiment of current tweets (anger, happy, love, fear, envy, surprise, sadness).

One of the first analyses of the relationship of the state of Twitter users and contemporaneous real world events was made by Bollen et al. (2009). They extracted collective mood from Twitter, more specifically they defined its six dimensions (calm, alert, sure, vital, kind, happy). Then, on a data sample from August 2008 to December 2008 in the USA, they studied the impact of various socio-economic events (changes in oil prices, stock markets prices, presidential elections, Thanksgiving) on individual dimensions of collective mood.

Similarly, Curini et al. (2015) analyzed the impact of various characteristics of Italian provinces on the overall happiness of Twitter users in 2012. They found that the current weather and the spread between German and

[4] http://www.ccs.neu.edu/home/amislove/twittermood/?utm_campaign=Facebook+Page&utm_content=Pulse+of+the+Nation+US+Mood+Throughout+the+Day+inferred+from+Twitter

Italian bonds had the highest impact, while relatively stable characteristics, such as the quality of institutions, had almost no impact.

Others have studies the possibility to use the information about the current state of Twitter users, together with a linkage to geographical location when tweeting, in real life applications, such as mapping the situation during catastrophes. Earle et al. (2010) took Twitter data from March 2009 in the USA to provide a detailed mapping of the impact of an earthquake – precise localization of the most affected areas as well as the severity of effects described by users. Guan & Chen (2014) did the same for the hurricane Sandy of 2012.

### 2.4.2 Forecasts outside of economics

The first analysis of using Twitter data to forecast the outcome of real world events was conducted by Asur & Huberman (2010), namely box office revenues in the USA over the period from November 2009 to February 2010. Taking tweets that mentioned twenty four films released over this period, they calculated a so called ”tweet-rate” (number of tweets containing the name of the film per hour) and also the sentiment of such tweets (negative, neutral, positive).[5]

They included their tweet-rate as an explanatory variable when modeling box office revenues of individual films, finding that their measure performed well compared both with Hollywood Stock Exchange index and a model using Internet Movie Database and news feed data. They also found that incorporating the index of sentiment (number of positive tweets divided by number of negative tweets) further improved their predictions, although the tweet-rate remained dominant.

Similarly to Google data, Twitter data have been analyzed in various scientific fields, often overlapping with those of Google data application. For example Culotta (2010), similarly to Ginsberg et al. (2009), predicted influenza occurrence in the USA on a ten weeks data sample from February 2010 to April 2010. Other areas of application include social sciences, such as politology. Tumasjan et al. (2010) analyzed tweets from Germany one week before the elections in September 2009, finding that a mere number of mentions of a political party reflected well upcoming election results. Ko et al. (2014) studied the same for the presidential elections in South Korea in 2012.

[5] The films were chosen such that their name was not interchangeable with other potential uses of similar text strings unrelated to the film.

The difference between Google and Twitter data is, of course, the nature of the input. While Google query most often contains two or three words,[6] tweets may have up to 140 characters in total. And while Google data is available processed and can be taken either as a search volume index for individual words or for whole categories, researchers using Twitter data must usually conduct a linguistic analysis to assess which tweets correspond to a phenomenon they aim to examine. That could be tweets containing a particular string or a specified set of words etc.; analyzing only single words would often introduce large noise caused by the inclusion of irrelevant tweets.

### 2.4.3 Forecasts in economics

In their paper, Mao et al. (2011) aimed to compare three sources of on-line data defined above. On a dataset from July 2010 to September 2011, they took survey data (as a traditional source), news headlines, search engine data, and Twitter data, and analyzed their interconnection with financial variables (Dow Jones Industrial Average price, trading volumes, market volatility VIX, and gold prices). They used survey data, news headlines and Twitter data to measure sentiment of investors (e.g. Twitter sentiment as a share of tweets containing the word "bullish" relative to "bearish"), and Google data and Twitter data to measure the volume of search or the number of mentions in tweets for individual financial terms.

When predicting individual time series of financial variables, they found that while Google data performed well, survey data were not useful. Further, both Twitter data and news headlines data proved to be a significant predictors for Dow Jones Industrial Average and VIX. The authors also concluded that Twitter data may be more useful than Google data, since they perceived a more timely development of Twitter indices compared with other tested variables.

Following their extraction of collective mood from Twitter, Bollen et al. (2011) also analyzed its correlation with financial markets, specifically with Dow Jones Industrial Average on a data sample from February to December 2008. They found that the forecasts of DJIA may be significantly improved when using some dimension of collective mood, while other dimensions were not useful. Other authors, such as Zhang et al. (2011) or Porshev et al. (2013), also studied the potential to use Twitter sentiment data to improve

[6] http://blueniileresearch.com/psychology-searcher/

forecasts of US financial market indicators, finding models with Twitter data to be superior.

Souza et al. (2015) then modeled the interrelation between Twitter mentions (volume as well as sentiment) of five listed US retail brands listed and their stock returns and volatility in the period from November 2013 to September 2014. Using also news data (Wall Street Journal, Dow Jones Newswire) as control variables, they found that Twitter data bring useful additional information.

Papaioannou et al. (2013) were also concerned with financial markets, more specifically the EUR/USD exchange rate and the possibility of its predictions. On a data sample from October 2010 to January 2011, they measured it on a high-frequency, intra-day trading scale. In their case, they used a specific kind of tweets containing a string "buy EUR/USD" followed by a specific target price – these tweets were posted either by individual investors or by on-line algorithm brokerage firms that publish incoming limit orders of their retail clients on Twitter.

The authors view such limit orders as beliefs of individual investors about the development of the exchange rate and aimed to discover whether these investors – on average – knew more than the market. Aggregating the data on an hourly basis, the authors found that Twitter data didn't bring any significant improvement of forecasts (measured by RMSE and MAE) compared to random walk, but improved the statistics concerning the number of ups and downs correctly predicted. Most notably, while a trading strategy based on the random walk assumption was loss-making, a strategy using Twitter data proved profitable. The authors concluded that at least in some cases – in a very short-term horizon – the forex markets are ineffective as the prices do not include all available information. On the other hand, the authors note that this could be caused by a trend contained in a short term, as the results did not confirm for longer horizons.

Janetzko (2014) also analyzed the possibility to predict the EUR/USD exchange rate, but chose a completely different approach. In their understanding, Twitter is a global and fast aggregator of news and opinions, and an increase of some particular terms means an increased "attention to some topic relative to others". On a data sample from January 2012 to September 2013, they analyzed the set of words connected to the Eurozone crisis and Greek debt crisis, and anticipated that an increased attention to this topic should translate to the depreciation of Euro. They indeed found that using tweets count (tweets containing a particular words, namely tweets containing Euro & crisis & SP and Euro & crisis & risk), the EUR/USD exchange rate "proved forecastable above chance level". Therefore, despite a vastly different methodology applied, the authors' conclusion was similar to Papaioannou et al. (2013).

Also concerned with the Eurozone and Greek debt crisis were Dergiades et al. (2013). They analyzed the influence of social media activity (Twitter, Facebook) and web search queries (Google) on a sovereign spread between GIIPS countries (Greece, Italy, Ireland, Portugal, Spain) and German long-term government bond yield. On a data sample from May 2011 to May 2013, they found that the discussion in social media and web searches connected to the Greek debt crisis contained significant information about the development of spread of debt financing costs between Germany and peripheral Eurozone countries, even when controlling for financial variables (idiosyncratic default risk, liquidity risk, international risk).

Most notably, the authors found that there was a one-way short-run causality in the direction from social media to the Greek spread, but not the other way around (neither short-term nor long-term), the data about Greek debt crisis from Twitter were also useful when explaining the development of Portuguese and Italian spread (indicating a weak contagion effect); and concerning the high-frequency data, Twitter and Facebook data proved better compared to Google, indicating that Twitter became a popular means of spreading news and news analysis.

Among applications most relevant to this thesis, Antenucci et al. (2014) studied Twitter data in the case of labor market flows. On a data sample from July 2011 to November 2013 in the USA, they tried to predict the development of the official unemployment claims statistics by analyzing the volume of tweets indicating the user lost his job. Analyzing tweets containing strings relevant to job loss, and aggregating individual time series using principal component analysis, they created an index that proved to have a significant relationship with the official statistics.[7] On the other hand, the data didn't only copy well the official statistics, but also brought an independent own information that could be explained by anecdotal evidence. For example a case of a system failure of official statistics that was recovered the next week (measurement error), and Twitter data remaining stable.

When creating forecasts, they found that while Twitter data were not as good a predictor as lagged values of new claims and consensus data, it still remained a significant explanatory variable in a model containing the previ-

[7] The strings analyzed included for example "lost job", "fired", "canned", "I lost my job", "laid off", "unemployment" and others, with strings up to 4 words.

ously mentioned ones as control variables. The authors were encouraged by the performance of Twitter data that they created a recursively updated publicly available Michigan Social Media Job Loss Index.[8] The authors also used metadata contained in tweets (users mentioning their age, sex etc.) to make profiles of a sample of users. They found that the distribution of Twitter users is not identical to the population level, and that concerning the communication about job losses, older people are overrepresented relative to their share on Twitter.

[8] http://econprediction.eecs.umich.edu

# 3 Data

## 3.1 Google data

### 3.1.1 General description

Google provides the data about search volumes at Google Trends website (www.google.com/trends). The information in this subsection is based also on the official description of the service.[9] In general, Google provides the share of searches for a particular query to the total number of searches conducted through Google search during a given period in a given location. Google started to publish the data in a complete form in 2009, all time series are available starting in the first week of 2004. The data are of a weekly frequency and are updated almost immediately by the end of each week.

The final value for each query is normalized in several ways. Firstly, it measures the popularity of a particular query, not the absolute volume of searches. This serves better to compare the popularity of the query both during time and between countries. The final values of a time series are further normalized to be between 0 and 100 (maximum over the period). The value of zero is assigned also when the absolute number of searches for that query was below a particular threshold (not officially specified). Also, queries are excluded from the analysis if they are searched for too little; and when they are searched for by the same user repeatedly in a short period of time, such data do not enter the statistic.

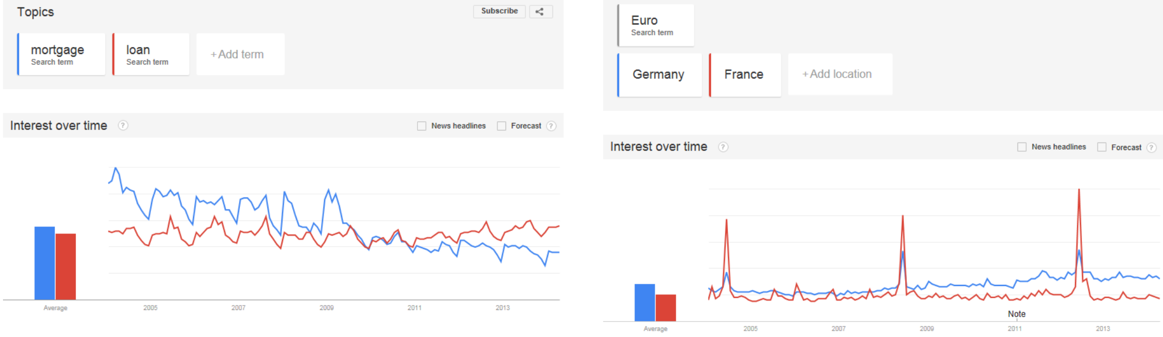
It is possible to specify a time period over which we wish to study the volume of search; a geographical location of interest where the query was searched for; and for some languages, words are divided into categories and sub-categories based on their meaning. It is possible to compare the popularity of search between more queries (up to 5) or to compare the popularity of some query in different locations.

Figure (3.1) depicts such comparisons. The left-hand-side (LHS) shows the comparison of a popularity of search queries ”mortgage” and ”loan” worldwide from 2004 to 2013; the relative volume of searches for ”mortgage” declined sharply in the first quarter of 2009 and the popularity of ”loan” has

[9] https://support.google.com/trends/answer/4355213?hl=en&ref_topic=4365599

been slowly rising. The right-hand-side (RHS) chart compares the popularity of search for query ”Euro” in France and Germany.

**Figure 3.1: Search volume data – comparison of queries and geographical locations**




Queries "mortgage" and "loan", worldwide. Query "Euro" in France and Germany.

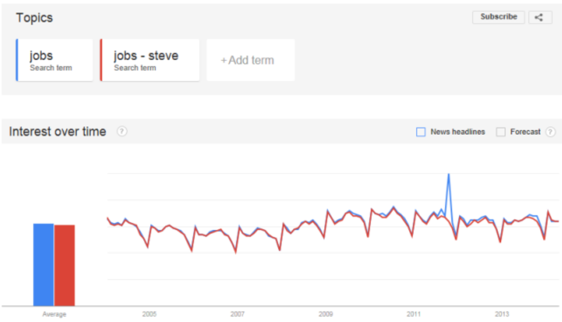
Source: Google Trends

Sharp peaks are clearly visible over the period 2004–2013, happening four years apart in both countries coincidentally (even though peaks are higher in France) – this shows a one-time interests in the UEFA European Championship in football every four years rather than the currency of Euro. This demonstrates two of the threats of using such data – firstly, researchers have to consider in what context a word can be used (or what kind of noise it may include); secondly, a one-off events may influence the data severely.

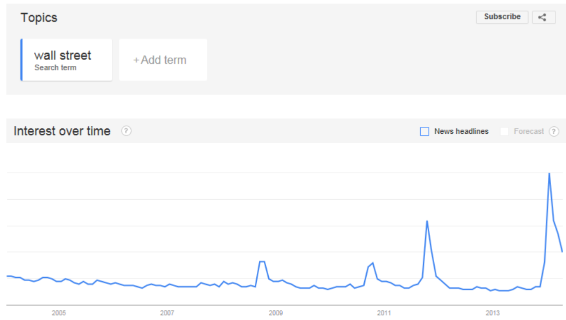
Researchers may adjust the analyzed search query with an aim to include all wanted information and remove some noise. It is possible to clean the data of a particular search query from queries that included also other words that would indicate that the user was interested in something different. For example, by entering a query ”jobs”, a person can be interested in the topic of employment, but also in information about Steve Jobs. It is possible to analyze a string ”jobs - steve” which gives the data for all queries containing the word ”jobs” that did not contain the word ”steve”.

The LHS of Figure (3.2) shows the chart for this exercise worldwide. There is no systematic difference between these two series, only a one-off impact of people searching for information about Steve Jobs in October 2011 around the time of his death. For this reason, D’Amuri & Marccuci (2012) cleaned the data from this effect before they used them to analyze unemployment. Other queries may be affected more systematically when the phenomenon causing the noise is searched for during the whole analyzed period. For every query, Google suggests a list of related ones to show in what context people searched for it the most; this serves as an indication to researchers what kind of noise may be included.

**Figure 3.2: Search volume data – examples of cleaning and compounding**




Queries "job" and "job - steve", worldwide.




Query "wall street", worldwide.

Source: Google Trends

Instead of deducting words, it is also possible to add them together (using ”+” sign) to see the aggregated time series of volume of search of queries that contained one of the words (possibly up to 25 elements). People also search for queries containing more than one word, the RHS of Figure (3.2) shows this for a query ”wall street” (should be in parentheses if we are interested in the exact order of words, or without if we only care about all words being contained in a search query). Peaks show one-off events such as the protest group Occupy Wall Street in October 2011 or the film The Wolf of Wall Street in January 2014.

### 3.1.2 Data availability – categories, time periods, sampling, and download

For some languages, Google also divides queries into categories. It is then possible to see the time series of a volume of search for all queries in a given category (relative to all searches conducted), such as ”Jobs & Education” or ”Internet & Telecom”. This serves as a source of information instead of using individual queries (if the category covers some phenomenon to be explained). The other use is to see the volume of search of a particular query based on the context of the search; often quoted is the case of ”apple” being either in the ”Food & Drink” or ”Computers & Electronics” category. Unfortunately, this feature is not available for the Czech language, so the analysis of search volumes can be done only using individual queries.

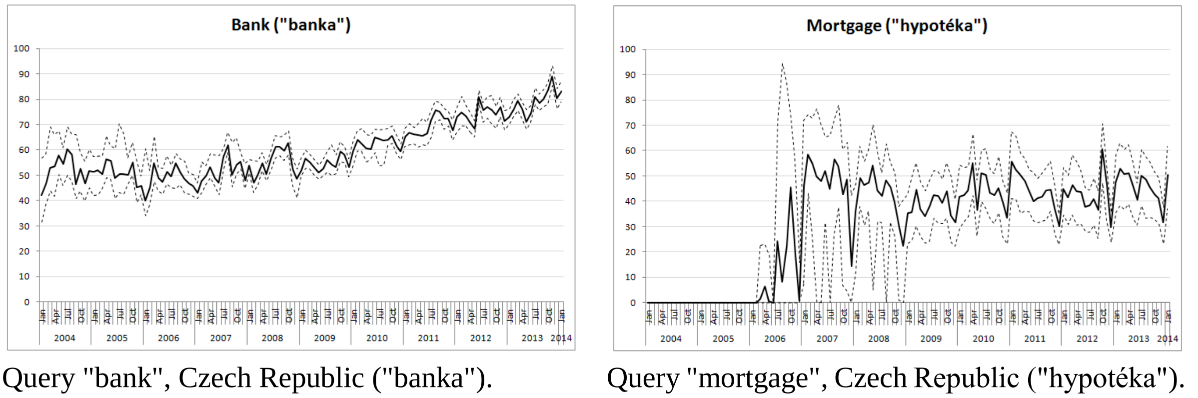
All time series start in the first week of January 2004, but it is possible to restrict the length of the series to see only development during particular period. Based on the choice of the period shown, the data frequency slightly differs. On the web pages interface, daily observations are shown for periods shorter than 3 months, weekly observations for periods shorter than 3 years, and monthly observations if the period is longer. Google also allows direct downloads of the data in CSV format (if the user is logged in into Google services) and in that case, weekly data can be downloaded even for periods longer than 3 years.

For the purpose of our thesis, we downloaded weekly data from January 2004 to December 2013, restricting the geographical location to the Czech Republic. We then transformed the data into monthly frequency: weekly data were assigned to each day of that week, and an arithmetic average was calculated over all days of each month.

Even though Google does not specifically mention reason for this or even the fact in general, the data downloaded in different points in time slightly differ. Da et al. (2013) claim that "*to increase the response speed, Google currently calculates the search volume index from a random subset of the actual historical search data*". We could not find a direct confirmation of this on Google sites, but Choi & Varian (2012) confirm it by saying "*Google Trends data is computed using a sampling method, and the results therefore vary a few percent from day to day.*"; indeed, the time series differ even when downloaded 24 hours apart (but not less).

For this reason, data for all Google queries used in our analysis were downloaded every day during February 2014, a total of 28 times; an arithmetic average over these 28 time series was taken as a representative series for each of the queries.[10] Figure (3.3) shows how data downloaded on different days differed for two chosen queries – "bank" and "mortgage" ("banka" and "hypotéka" in their Czech translation).

**Figure 3.3: Search volume data – average, maximum and minimum values**




Query "bank", Czech Republic ("banka"). Query "mortgage", Czech Republic ("hypotéka").

Source: Google Trends, author's calculations

Explanation: The average (solid line) was calculated from 28 time series of this query downloaded on different days; the minimum and maximum values (dotted lines) show the minimum or maximum value from all 28 time series for each month.

[10] Section 8.4 discusses practical aspects of averaging the data downloaded on 28 consecutive days.

The figure shows the average value (solid line) and maximum and minimum values (dashed lines) for that particular month from 28 downloaded series. Both charts depict some of the characteristics of the search query data for the Czech Republic: some queries are relatively popular (the absolute volume of searches is sufficient), the time series is full and even the data downloaded on different days do not differ much (if they do, it is rather in the beginning of the sample). On the other hand, some queries return zero values for the first few years and are volatile even later on with a big gap between maximum and minimum values depending on the day of the download. Such series become relatively stable only by the end of the analyzed period.

## 3.2 Dependent and control variables

The following section briefly describes the set of variables used as dependent and explanatory variables in the following analysis in addition to Google data. Main source of the data is the Czech Statistical Office (CZSO), together with OECD and the Czech National Bank (CNB).

### 3.2.1 Unemployment rate (CZSO)

By the definition of CZSO, a person of age 15 and older has to satisfy the following conditions to be considered unemployed:

(1) not to be employed;
(2) actively look for a job;
(3) be prepared to enter a job position.

The unemployment rate is the share (expressed in percentages) of unemployed to the total work force (the sum of all employed and unemployed, also called economically active population). CZSO gathers the data about employed and unemployed people through Labor Force Survey. The survey and its main results are quarterly in their nature and are published with a lag of three months after the end of the reference quarter. These include for example division by geographical regions, age, gender, education, and the rate of long-term unemployment. Approximately one month after the end of the references month, CZSO publishes preliminary results that contain only the rate of unemployment for people of age 15-64 (broken down by gender), all the time series are seasonally adjusted.

### 3.2.2 Share of unemployed (Labor office)

Czech Ministry of Labor and Social Affairs publishes its own statistics of unemployment based on own sources of data. Since 2012, they publish so called "share of unemployed person", which is a share of people of age 15-64 registered at the Labour office ("Úřad práce") to the total population of that age. This series differs from the official unemployment rate of CZSO since the numerator contains only people registered as unemployed (which does not follow the general definition of unemployment) and the denominator contains whole population (not only economically active people). It is published approximately 10 days after the end of the reference month. Due to recent changes in methodology, the data are available beginning in January 2005.

### 3.2.3 GDP, national accounts (CZSO)

GDP is the ultimate measure of performance of the economy, and as CZSO defines "*it represents the sum of values added by all branches of activities which are considered productive in the system of national accounts (including market and non-market services). Calculations are made at current prices and results are then converted into constant prices so that development excluding influences due to price changes can be kept track of*".[11]

It is the total value of goods and services created in a given period in a given area expressed in monetary units. There are three basic approaches used for calculation: (a) production approach; (b) income approach; and (c) expenditure approach. The national accounts data are published with a quarterly frequency in several forms. With a delay of 1.5 months after the end of the reference quarter, preliminary estimates are published in a form of basic information (growth / decline of the total GDP in percents). After next three weeks, more complete data are available, but revised values are published with a delay of three month after the end of the reference quarter.

### 3.2.4 Confidence Indicators (CZSO)

Confidence indicators published by CZSO are based on business cycle surveys during which chosen individuals (businessmen, consumers) are asked about their expectations about future development in the economy. Questions asked in these surveys are qualitative, respondents are supposed to assess their expectations about the future (month, quarter, year) compared to current situa-

[11] http://www.czso.cz/eng/redakce.nsf/i/gross_domestic_product_(gdp)

tion (better, worse, the same). The goal is to provide information about the situation, mood, and atmosphere both on the business side of the economy as well as on the consumers' side, with the aim to identify the turning points in the economy.

There are three resulting indicators: Business Confidence Indicator (BCI), Consumer Confidence Indicator (CCI) and Composite Confidence Indicator (CI) which consists of BCI and CCI with weights 80% and 20%. They are published as basic indices approximately one week before the end of the reference month. More detailed methodology for Consumer Confidence Indicator is discussed in Chapter 6.

### 3.2.5 Composite Leading Indicators (OECD)

CLIs is an index published by OECD for 39 countries (33 member and 6 non-member countries), its description below is based on the official information from the methodological document OECD (2012). The purpose of this index is to give signals about the turning points in the economy from 4 to 8 months in advance. The index consists of more underlying indicators that display similar cycle and also lead the business cycle represented by a reference series; Index of Industrial production served as a reference series until 2012 and it is GDP since then.

There has to be both statistical as well as theoretical relationship of such indicator and the reference series, the frequency and publication delay also play an important role. The choice of component series is evaluated regularly and changes are often made. Currently OECD (2014) list 7 component series for the Czech Republic: Balance of payments Capital account, debit (czk); Demand evolution (Services): future tendency (% balance); Production (Manuf.): tendency (%); CPI Harmonised All items (2000=100) inverted; Consumer confidence indicator (% balance); ITS Exports f.o.b. total (czk); Share prices: PX-50 index (2010=100). CLIs is available with a monthly frequency with a delay of approximately 6 weeks after the end of the of the reference month.

### 3.2.6 Peaks and trough dates (OECD)

In the above mentioned source, OECD (2014) also provides a list of peaks and troughs for each of the countries for which the CLIs is published. These peaks and troughs are derived from the reference time series (which was Index of Industrial Production until 2012 and is GDP since then) and it can provide information about the current state of the economy in terms of growth or decline.

### 3.2.7 Other control variables

Other control variables acquired from the Czech Statistical Office are Index of Industrial Production, Average Wage, and two measures of inflation – Consumer Prices Index and Prices of Industrial Producers. Index of industrial production is a measure of output of the industry in total, as well as of individual industrial economic activities. The calculation combines two measures – sales of own goods and services in most economic activities, and physical production volumes in others. When aggregating the data from individual activities, their weights reflect their value added in the base year.

The statistics of Average Gross Monthly Wage, as provided by the CZSO, includes basic wages and salaries, supplementary payments, bonuses and other remunerations, but excludes other personnel expenses. The statistics takes into account the type of employment to provide figures on a full-time equivalent basis. Consumer Prices Index measures the evolution of the price level of a representative consumption basket which contains foodstuffs, other goods, and services, and where the weights of individual items reflect the share of spending on this item in total household consumption. Prices of Industrial Producers measure the average evolution of prices of all industrial products produced and sold in the Czech market, while it captures the agreed upon prices between the supplier and the customer (without VAT).

Time series downloaded from the website of the Czech National Bank (CNB) include monthly averages of the CZE/EUR exchange rate, volumes of monetary aggregates M1, M2 and M3, the Prague InterBank Offered Rate (PRIBOR), the volume of loans to households by commercial banks, and the interest rate on credit to households.[12] Lastly, an index of the Prague Stock Exchange – the PX Index – was used, specifically its last day of a month value was assigned to each month.

[12] http://www.cnb.cz/docs/ARADY/HTML/index.htm

# 4 Methodology

To examine the interconnection between Google data and real macroeconomic variables, we will use two of the classic models commonly applied in economic time series modeling: Autoregressive Integrated Moving Averages (ARIMA) and Vector Autoregression (VAR); also, logit model for dichotomous dependent variable will be used. Methodological information in this chapter is based mostly on Maddala (2001), Enders (2003) and Brooks (2008). It provides information about ARIMA (4.1), VAR (4.2) and logit models (4.4), readers familiar with individual models are encouraged to skip the respective section. In addition, forecasting (4.3) and seasonality issues (4.5) are discussed.

## 4.1 ARIMA, ARMAX

ARMA models belong into the group of univariate models. The current value of a variable is modeled using only information contained in its lagged values and in the lagged values of the error term. It consists of two parts: autoregressive (AR) and moving averages (MA).

An autoregressive process of order $p$ is defined:

$$\mathrm{AR(p)}: y_t = \alpha + a_1 y_{t-1} + \cdots + a_p y_{t-p} + \varepsilon_t = \alpha + \sum_{i=1}^{p} a_i\, y_{t-i} + \varepsilon_t \qquad (4.1)$$

where $a_i$ are coefficients, $y_{t-i}$ are lagged values of the dependent variable and $\varepsilon_t$ is white noise (a process with constant mean and variance, zero value of autocorrelation coefficients, and normal distribution).

Process of moving averages of order $q$ is defined:

$$\mathrm{MA(q)}: y_t = \beta + \varepsilon_t + b_1 \varepsilon_{t-1} + \ldots + b_q \varepsilon_{t-q} = \beta + \varepsilon_t + \sum_{i=1}^{q} b_i \varepsilon_{t-i} \qquad (4.2)$$

where $b_i$ are coefficients, $\varepsilon_t$ is white noise and its lagged values. Those are assumed to be independently identically distributed with mean equal to zero and variance equal to $\sigma^2$ [$\varepsilon_t \sim IID(0, \sigma^2)$].

When combined together, the result is an $\mathrm{ARMA(p,q)}$ process:

$$\mathrm{ARMA(p,q)}: y_t = \gamma + \sum_{i=1}^{p} a_i\, y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} b_i \varepsilon_{t-i} \qquad (4.3)$$

where the white noise term is $\varepsilon_t \sim IID(0, \sigma^2)$. In this overall process, the dependent variable is explained by a linear combination of its own lagged values and a combination of current and lagged values of the error term.

An economic theory can propose additional explanatory variables that can be added into the ARMA process. Those can be added as a set of current and lagged values (distributed lag model), or only as a particular lag; we define $\mathrm{ARMAX}(\mathrm{p},\mathrm{q})$ model:

$$\mathrm{ARMAX}(\mathrm{p},\mathrm{q})\colon y_t = \gamma + \sum_{i=1}^{p} a_i\, y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \sum_{i=1}^{r} c_i x_i \qquad (4.4)$$

where $x_i$ are additional explanatory variables, $r$ is their quantity and $c_i$ their coefficients. In this case, however, we move from univariate to multivariate models.

Statisticians G. Box and G. Jenkins contributed in the popularization of ARMA models by providing general procedure of model estimation. Box-Jenkins methodology consists of five steps:

(1) Checking and achieving stationarity.
(2) Model identification.
(3) Model estimation.
(4) Diagnostics and model checking.
(5) Forecasting.

### 4.1.1 Stationarity

A time series is said to be stationary if its probability distribution is not depended on time. (And a pair of time series is jointly stationary if its joint probability distribution is not dependent on time.) In its weak form, we recognize three conditions for a time series to be stationary.

(a) $E(y_t) = \mu$
(b) $E(y_t - \mu)(y_t - \mu) = \sigma^2 < \infty$
(c) $E\left(y_{t_1} - \mu\right)\left(y_{t_2} - \mu\right) = \gamma_{t_2 - t_1}\ \forall t_1, t_2$

That is the mean of such time series is constant, its variance is constant and finite, and the covariance between two time periods is dependent only on the distance between these periods.

When time series are not stationary, problems may occur when estimating models. One of them is spurious regressions – obtaining significant coefficients even when there is no relation between two variables – caused only by a common trend. In autoregressive models, on the other hand, estimates of coefficients may be biased towards zero. Also, assumptions about asymptotic distribution are not valid when using non-stationary data, which has an impact on hypothesis testing.

MA(q) process is always stationary, AR(1) process is stationary if $|a| < 1$ and an AR(p) process is stationary if all roots of its characteristic polynomial $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$ lie outside of the unit circle. If the series contains a unit root, there is a permanent effects of shocks; when $|a| < 1$, effects of shocks will disappear.

Both formal and informal tests exist to determine stationarity of a time series. Informal ones are based on observations of deviation from the three conditions of weak stationarity: the time series does not have a constant mean, variance, or it has a "long memory" – its autocorrelation function (ACF) declines only slowly with increasing distance.

There are two basic formal statistical tests for stationarity: Augmented Dickey-Fuller test (ADF) and Kwiatkowski-Phillips-Schmidt-Shin test (KPSS). ADF test analyzes whether the time series contains a unit root, its null hypothesis is non-stationarity. On the other hand, KPSS's null hypothesis is stationarity of the series.

To achieve stationarity, the series can be transformed by differencing. If the series become stationary after first differencing, we call it integrated of order one; if we need to difference it twice, it is integrated of order two and so on. When we apply $\mathrm{ARMA(p, q)}$ on such differenced series, it is called $\mathrm{ARIMA(p, d, q)}$, where $d$ is the order of integration. Since a lot of economic time series are not stationary, imposing such transformation is common.

As Stock & Watson (2007) note, logarithmic differences are often applied to economic time series, because a lot of them grow approximately exponentially and logarithm of such series thus grow in a linear fashion. Other reason is that standard deviation of many economic series is approximately proportional to their level. In that case, standard deviation of a logarithm of such series is approximately constant. In both cases, it is useful to transform the original series in such a way that the changes in transformed series would be proportional to changes in the original one, and logarithmic differences achieve this.

### 4.1.2 Model Identification

Next, correct orders $p$ and $q$ have to be determined. This can be done through graphical analysis of regularities in the plot of the autocorrelation function (ACF) and partial autocorrelation function (PACF). But since the observation of patterns is usually complicated with real economic data, information criteria are used instead. Information criteria take into account residual sum of squares with the aim to minimize it, but also penalize the loss of degrees of freedom caused by estimating additional parameters.

There are three commonly used information criteria: Akaike (AIC), Schwartz (BIC) and Hannan-Quinn (HQIC), they can be defined as follows.

$$\begin{aligned} \mathrm{AIC} &= \ln(\hat{\sigma}^2) + \frac{2k}{T} \\ \mathrm{BIC} &= \ln(\hat{\sigma}^2) + \frac{k}{T}\ln T \\ \mathrm{HQIC} &= \ln(\hat{\sigma}^2) + \frac{2k}{T}\ln(\ln T) \end{aligned} \tag{4.5}$$

where $\hat{\sigma}^2$ is residual variance (residual sum of squares divided by the number of observations) and $k = p + q + 1$, total number of parameters to estimate. The goal is to minimize these criteria, calculating them for a set of models with predetermined thresholds $p \leq \bar{p}$ and $q \leq \bar{q}$. Residual sum of squares declines with adding more lags into the model, but the loss of degrees of freedom is penalized; penalization is the biggest in the case of BIC and the lowest with AIC; HQIC stands in-between.

In general, BIC is consistent (chooses the correct specification asymptotically); AIC is not consistent (on average chooses bigger than correctly specified model) but is more efficient than BIC. None of the criteria is generally better than the other, but some recommendations exist about their choice for different time series (e.g. by frequency). Information criteria can be used also when estimating models with additional explanatory variables, but those are defined slightly differently.

### 4.1.3 Model estimation

In this step, we estimate coefficients of the model specified in step 2. Depending on the model, the estimations are made either using the method of least squares or maximum likelihood.

### 4.1.4 Diagnostics and model checking

The goal of ARMA modeling is to create a parsimonious model that captures all significant relations in the data. In this step, we look for possible signs of overfitting the model (using larger specifications than necessary to capture the dynamics), as well as testing residuals for possible preserved linear dependencies. This can be done again using ACF and PACF as in step (2), but also with a Ljung-Box test.

Ljung-Box test is a test of preserved linear dependencies in residuals, and serves as a general portmanteau test. It tests the joint null hypothesis that the first $m$ values of autocorrelation function are equal to zero. An alternative hypothesis is at least one of these values being non-zero. If some dependencies remained in residuals, we should get back to step (2) and specify the model correctly.

### 4.1.5 Forecasting

According to Enders (2003), one of the most important uses of ARMA models is for forecasting future values of the time series in question. Therefore, comparing predictive capabilities of competing models is the best way how to determine their quality. Such forecast are made either one-step ahead, where coefficients estimated over a sample ending one period are used directly to forecast values of the dependent variable for the next period, or also for two- or more steps ahead.

For example, let us have an ARMA(2,2) model:

$$y_t = \gamma + a_1 y_{t-1} + a_2 y_{t-2} + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \varepsilon_t \tag{4.6}$$

firstly, the model is estimated on a given sample to obtain estimates of all parameters $\hat{\gamma}$, $\hat{a}_1$, $\hat{a}_2$, $\hat{b}_1$, $\hat{b}_2$ as well as the residual $\hat{\varepsilon}_t$; past values of dependent variable are known and past values of the error term were estimated in previous steps. To forecast the value for the next period, these estimates are plugged into the equation (4.6):

$$\hat{y}_{t+1} = \hat{\gamma} + \hat{a}_1 y_t + \hat{a}_2 y_{t-1} + \hat{b}_1 \hat{\varepsilon}_t + \hat{b}_2 \hat{\varepsilon}_{t-1} + \hat{\varepsilon}_{t+1} \tag{4.7}$$

The only unknown value on the right hand side is $\hat{\varepsilon}_{t+1}$ which will be replaced by zero, its expected value. Similarly, we can continue with forecast more steps ahead, only the values of $y$ on the right hand side would not be known and their predictions would be used (so called dynamic forecast).

As Brooks (2008) mentions, ARIMA models often proved to be better predictors in a short horizon compared with structural economic models, but were less precise for longer horizon and were not able to capture unusual changes in the dependent variable. Therefore, additional explanatory variable can help to add new dynamics into the model. In that case, the forecasting procedure looks the same – the right hand side only contains more coefficients to estimate.

## 4.2 VAR

When using additional explanatory variables, a question about their exogeneity can be raised – that they should be treated more symmetrically as endogenous variables. This can be achieved using simultaneous equations model, but it also requires dividing variables into exogenous and endogenous in advance, and imposing further restrictions to achieve identification.

Sims (1980) criticized this procedure because a researcher has to make arbitrary decisions in almost every step. He proposed an alternative in a form of multiple time series generalization of AR models and as a consequence, he popularized the use of Vector autoregression (VAR) in econometrics as an extension of univariate AR models.

In that case, there is a system of interconnected variables where the current value of each variable is determined by own past values and also by current and past values of all other variables. A bivariate VAR system with up to $p$ lags for each variable is defined as follows:

$$\begin{aligned} y_{1t} &= b_{10} + b_{12}y_{2t} + \textstyle\sum_{i=1}^{p}\gamma_{11}^{i}y_{1,t-i} + \sum_{i=1}^{p}\gamma_{12}^{i}y_{2,t-i} + \varepsilon_{1t} \\ y_{2t} &= b_{20} + b_{21}y_{1t} + \textstyle\sum_{i=1}^{p}\gamma_{21}^{i}y_{1,t-i} + \sum_{i=1}^{p}\gamma_{22}^{i}y_{2,t-i} + \varepsilon_{2t} \end{aligned} \tag{4.8}$$

White noise disturbances $\varepsilon_{it}$ are uncorrelated ($E[\varepsilon_{it}] = 0, E[\varepsilon_{1t}, \varepsilon_{2t}] = 0$), but because of the interconnection of both variables in the system, correlation exists between these variables and error terms ($cov[y_{1t}, \varepsilon_{2t}] \neq 0, cov[y_{2t}, \varepsilon_{1t}] \neq 0$) and for this reason, standard estimation techniques cannot be used, since Gauss-Markov theorem is violated.

The expression (4.8) is so called structural form of the VAR system, it can be rewritten into a matrix form:

$$\begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix}\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \textstyle\sum_{i=1}^{p}\begin{bmatrix} \gamma_{11}^{i} & \gamma_{12}^{i} \\ \gamma_{21}^{i} & \gamma_{22}^{i} \end{bmatrix}\begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{4.9}$$

and when denoting:

$$B = \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix}, y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix}, \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix}, \Gamma_i = \begin{bmatrix} \gamma_{11}^{i} & \gamma_{12}^{i} \\ \gamma_{21}^{i} & \gamma_{22}^{i} \end{bmatrix}, y_{t-i} = \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \end{bmatrix}, \varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

the system of equations is:

$$By_t = \Gamma_0 + \textstyle\sum_{i=1}^{p}\Gamma_i y_{t-i} + \varepsilon_t \tag{4.10}$$

If $B$ is invertible, pre-multiplying both sides of the equation by $B^{-1}$:

$$y_t = B^{-1}\Gamma_0 + \textstyle\sum_{i=1}^{p}B^{-1}\Gamma_i y_{t-i} + B^{-1}\varepsilon_t \tag{4.11}$$

and when denoting:

$$A_0 = B^{-1}\Gamma_0,\ A_i = B^{-1}\Gamma_i,\ e_t = B^{-1}\varepsilon_t$$

the final form is:

$$y_t = A_0 + \sum_{i=1}^{p} A_i y_{t-i} + e_t \tag{4.12}$$

This is so called reduced form of a VAR model and it is basically an autoregressive process $AR(p)$ for two (or generally more) variables at once. Also, when denoting:

$$A_0 = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix}, A_i = \begin{bmatrix} a_{11}^i & a_{12}^i \\ a_{21}^i & a_{22}^i \end{bmatrix}, e_t = \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

both equation can be expressed individually:

$$\begin{aligned} y_{1t} &= a_{10} + \sum_{i=1}^{p} a_{11}^i y_{1,t-i} + \sum_{i=1}^{p} a_{12}^i y_{2,t-i} + e_{1t} \\ y_{2t} &= a_{20} + \sum_{i=1}^{p} a_{21}^i y_{1,t-i} + \sum_{i=1}^{p} a_{22}^i y_{2,t-i} + e_{2t} \end{aligned} \tag{4.13}$$

Each of the two equations (4.13) can be estimated using ordinary least squares (OLS). Because cross-correlation is taken into account in the model, both time series need not be stationary, but the system has to be stable as a whole. The system is stable if all roots of inverse characteristic polynomial lie outside of the unit circle. However, time series should not contain a trend component. If they do, trend should be included in the system as an exogenous variable.

Some authors, including Sims (1980) himself, recommend not differencing the data to achieve stationarity, because the purpose of VAR is not to get precise estimates of coefficients, but to find all relations between variables, and differencing the data removes the information about long-term relationship. But to be able to test individual or joint hypothesis in the VAR system, such as the significance of coefficients, all component time series should be stationary.

To identify the correct lag order of the system, information criteria are used. They are analogous to the case of ARMA models, their definition slightly differs:

$$\begin{aligned} \mathrm{MAIC} &= \ln\left|\hat{\Sigma}\right| + \frac{2k'}{T} \\ \mathrm{MBIC} &= \ln\left|\hat{\Sigma}\right| + \frac{k'}{T}\ln(T) \\ \mathrm{MHQIC} &= \ln\left|\hat{\Sigma}\right| + \frac{2k'}{T}\ln(\ln(T)) \end{aligned} \tag{4.14}$$

where $T$ is the number of observations, $\hat{\Sigma}$ is variance covariance matrix of residuals defined as $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{22} & \sigma_2^2 \end{bmatrix}$, $\sigma_i^2 = var(e_{it})$, $\sigma_{12} = \sigma_{21} = cov(e_{1t}, e_{2t})$. And $k'$ is the total number of regressors in all equations; generally, for $k$ equations with a constant and $p$ lags, the total number of regressors is $k^2p + k$.

Recommendations exist as to what information criterion should be used. AIC is recommended for daily and monthly data, HQIC for quarterly data and BIC for quarterly data, yearly data and small data samples. Information criteria are used because when relying on classic F-test of significance of coefficients in individual equations, suggested number of lags may differ for individual equations. Following this suggestion would mean imposing restrictions, which is against the spirit of VAR modeling. Another possibility is a likelihood ratio test which compares variance covariance matrices of restricted and unrestricted models.

### 4.2.1 Granger causality

When estimating large VAR models, interpretation of results is difficult with big number of coefficients to analyze. For this reason, further tests and graphical representation of results were developed. One group of these tests examines a block significance, testing a hypothesis that a group of coefficients is jointly equal to zero, a special case of these tests is so called Granger causality test. Repeating the equations (4.13):

$$\begin{aligned} y_{1t} &= a_{10} + \textstyle\sum_{i=1}^{p} a_{11}^{i} y_{1,t-i} + \sum_{i=1}^{p} a_{12}^{i} y_{2,t-i} + e_{1t} \\ y_{2t} &= a_{20} + \textstyle\sum_{i=1}^{p} a_{21}^{i} y_{1,t-i} + \sum_{i=1}^{p} a_{22}^{i} y_{2,t-i} + e_{2t} \end{aligned} \tag{4.13}$$

The essence of Granger causality is testing a joint null hypothesis that for example all coefficients $a_{12}^{i}$ are equal to zero. If variable $y_2$ causes changes in variable $y_1$, at least some of the coefficients $a_{12}^{i}$ should by non-zero (and vice versa). If we reject the null hypothesis of zero values of coefficients, we say that "variable $y_2$ Granger causes variable $y_1$". This relation may exist in one or both direction, or not at all. It is not a causality in the sense of "cause and effect", rather a statistical dependence in the studied data sample.

### 4.2.2 Advantages and disadvantages of VAR

There are several points of critiques. First, a problem of data mining may occur because there is no theory in behind of a VAR model, for example concerning the selection of appropriate variables. More importantly, the number of parameters to estimate is large. For $k$ variables in the system with constant and $p$ lags of each variable in each equation, the total number of coefficients is $k^2p + k$ (for example for VAR(2) model for 3 variables, it is 27 coefficients). This causes a problem especially for short time series, degrees of freedom are used quickly.

On the other hand, VAR does not require researcher to make as many arbitrary decisions in advanced compared with simultaneous equations model (SEM), it is not necessary to determine which variables are exogenous and endogenous. Compared with AR models, it is possible to capture more relationships among variables, since the current values do not depend only on its past values, but also on the values of other variables.

Also in forecasting, the results of VAR models are often better than SEM or large structural models based on economic theory. It is because lagged values contained in VAR can overcome the misspecification error problem. Making forecast using VAR models is analogous to the case of ARMA. Having the reduced form:

$$\begin{aligned} y_{1t} &= a_{10} + \textstyle\sum_{i=1}^{p} a_{11}^{i} y_{1,t-i} + \sum_{i=1}^{p} a_{12}^{i} y_{2,t-i} + e_{1t} \\ y_{2t} &= a_{20} + \textstyle\sum_{i=1}^{p} a_{21}^{i} y_{1,t-i} + \sum_{i=1}^{p} a_{22}^{i} y_{2,t-i} + e_{2t} \end{aligned} \tag{4.13}$$

When we are concerned with the variable $y_{1t}$, analogously to ARMA forecasting, the first equation is estimated. One step-ahead forecasts are then calculated as follows:

$$\hat{y}_{1,t+1} = \hat{a}_{10} + \textstyle\sum_{i=0}^{p-1} \hat{a}_{11}^{i} y_{1,t-i} + \sum_{i=0}^{p-1} \hat{a}_{12}^{i} y_{2,t-i} + \hat{e}_{1t} \tag{4.18}$$

because all values on the right hand side are known or estimated and $\hat{e}_{1t}$ is replaced by its expected value equal of zero. For more steps-ahead forecasts, the process becomes more complex – one-step-ahead forecast are estimated for all variables in the system, and these predictions are used in the equation instead of known values.

## 4.3 Forecasting

Forecasting is one of the most interesting features of econometric models, as they can be used in real life. And as Brooks (2008) says, some econometrists even think that questions about coefficients significance or meeting assumption of classic linear regression models are irrelevant if the model produces precise predictions. From this point of view, evaluating the predictive power of a model is an important test of adequacy of the model. Also according to Clark & McCracken (2011), evaluating forecast performance has become a vital part of empirical time series analysis.

### 4.3.1 In-sample and out-of-sample division

There are two basic types of forecasts: in-sample and out-of-sample. In-sample forecasts are generated for the sample over which model parameters were estimated. But because it was the point of parameters estimation for the model to fit well over such sample, it is better to divide the whole data sample into two parts in advance: to estimate model parameters using the first part, and to conduct a so called pseudo out-of-sample forecasts for the second part of the sample.

They are called "pseudo out-of-sample" because even though they simulate the real life practice of forecasting, it is not a real out-of-sample forecasting since the whole sample is actually known in advance. This procedure provides a possibility to test the fit of the model even on data over which it was not estimated, and it also provides a way how to compare competing models that appear of equal quality using in-sample fit.

The choice of proportion in which to divide the data sample into in-sample and out-of-sample part is not trivial, it often depends on the discretion of an econometrist and various authors provide different recommendations. Generally, there is a sample of length $T$ divided into in-sample part of length $R$ and out-of-sample part of length $P$, so that $T = R + P$. Then, there can be up to $P - \tau + 1$ out-of-sample forecasts, where $\tau$ is the horizon of the forecast.

As Clark & McCracken (2011) note, a basic trade-off exists when choosing the length of $R$ and $P$. Bigger $R$ means using more observations for the model estimation, which should lead to more precise parameter estimates, and should, in theory, translate into more precise forecasts if the model was chosen correctly. On the other hand, bigger $P$ means more observations to assess the actual predictive qualities of the model, and tests for comparing different models have a better power.

For this reason, Clark & McCracken (2011) propose a simple rule of thumb to choose $P$ and $R$ such that $P/R$ is at least 1, or even higher, since the goal is to maximize the power of tests of predictive quality. For the same reason, Hansen & Timmermann (2012) also suggest the division to be made relatively early in the data sample to get as long out-of-sample part as possible. On the other hand, West (2006) proposes the ration of $P/R$ to be small, for example smaller than 0.1, so that the assumption of asymptotic irrelevance of errors of parameter estimates can be used.

Therefore, no generally accepted rule exists. As Hansen & Timmermann (2012) note, some authors take as little observations for in-sample estimation as possible and use the rest to compare forecast, others choose particular length of out-of-sample part and use the rest to estimate parameters; for example Stock & Watson (2007) propose simply 10% or 15% from the end of the sample. For ARMA models, Enders (2003) also provides a simple rule of thumb saying "*forecasts from an ARMA model should never be trusted if the model is estimated with fewer than 50 observations*".

As mentioned above, out-of-sample forecasts usually conducted are called "pseudo" because we do not have to wait for new realizations of the dependent variable. As Clark & McCracken (2011) point out, this could be a problem for many macroeconomic variables, because these are not only often announced with a delay, but are usually revised later (such as GDP values). Therefore, in a real life exercise, the data of a researcher would differ from those that are available to an econometrist who happens to be dividing a complete data sample.

### 4.3.2 Forecasting scheme

To allow evaluation of forecast, there should be a time series of forecasts for the same horizon (e.g. one-step-ahead, etc.). There are three basic schemes when conducting such exercise: fixed, rolling and recursive windows. (a) In the fixed window scheme, the model is estimated only once using first $R$ observations. These original parameter estimates are then used to make $P - \tau + 1$ out-of-sample forecasts, always plugging only new realizations of the dependent and explanatory variables in the estimated model. (b) With the rolling window scheme, the model is repeatedly estimated using $R$ observations, and the window moves along the data sample with each prediction made. Therefore, observations from the beginning of the series are gradually not taken into account when the model is estimated.

(c) On the other hand, recursive window scheme (or recursively expanding) takes into account the whole data sample. In the first step, first $R$ observations are used to estimate the model and a forecast is made. In the second step, first $R+1$ observations are used for the model estimate and again, a forecast is made, and so on. The purpose is to use all available information, and as Hansen & Timmermann (2012) note, it is the most efficient use of the data, even if it is often a source of heteroskedasticity.

### 4.3.3 Forecast evaluation

The first step in forecast evaluation is the calculation of forecast errors, which is the difference between the actual realization of the dependent variable and its forecasted value; the forecast error of model $i$ in period $t$ is:

$$e_{it} = y_t - \hat{y}_{it} \tag{4.19}$$

where $\hat{y}_{it}$ is the forecasted value.

Mean squared error (MSE), the most common measure of forecast accuracy, is defined:

$$\mathrm{MSE_i} = \frac{1}{n}\sum_{t=1}^{n} e_{it}^2 \tag{4.20}$$

where $n$ is the length of out-of-sample forecast (equal to $P-\tau+1$ in accordance with previous notation). Sometimes, its square root is used to get Root mean squared error (RMSE):

$$\mathrm{RMSE_i} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_{it}^2} \tag{4.21}$$

An alternative is Mean absolute error (MAE):

$$\mathrm{MAE_i} = \frac{1}{n}\sum_{t=1}^{n}|e_{it}| \tag{4.22}$$

In comparison with this, mean squared error is a quadratic loss function – it penalizes larger errors disproportionally more than smaller errors. MSE will by employed in our analysis, also in accordance with compared studies.

When comparing competing models, MSE statistics alone is not enough to state if one model's forecasts are significantly more precise; some sort of statistical test has to be conducted. In our analysis, two usually proposed test will be used: Diebold-Mariano for comparing non-nested models and Clark-

West for comparing nested models. In this context, two models are nested if the regressors in one model (the smaller one) are a subset of regressors in the other (bigger) model, so that the bigger model can collapse to the smaller one if particular coefficients are set to zero.

#### 4.3.3.1 Diebold-Mariano tests and its modification

In the early 1990s, Diebold & Mariano (1995) proposed a relatively simple test for comparing quality of two forecasts, and it has spread to common use since then; the description is based also on an article of Diebold (2012). In their interpretation, the comparison is model-free; this means that we look only at forecasts – or forecasting errors – forgetting the models that were used for their creation (in general, these models may not be even known). The forecasting error is then entered into a loss function $L(\cdot)$ that may in general have different forms, such as $L(e_{it}) = e_{it}^2$ (like MSE) or $L(e_{it}) = |e_{it}|$ (like MAE), etc. When comparing forecasts (of models) 1 and 2, the next step is to create a loss differential:

$$d_t = L(e_{1t}) - L(e_{2t}) \tag{4.23}$$

which, in our case, will have a form based on the Mean Squared Error:

$$d_t = e_{1t}^2 - e_{2t}^2 \tag{4.24}$$

The only assumption is that the loss function is covariance stationary. The null hypothesis is that the predictive accuracy of both models is equal, that is $E(d_t) = 0$. The test statistic is of a form:

$$\mathrm{DM} = \frac{\bar{d}}{\hat{\sigma}_d} \to N(0,1) \tag{4.25}$$

where

$$\bar{d} = \frac{1}{n}\sum_{t=1}^{n} d_t \tag{4.26}$$

and $\hat{\sigma}_d$ is a consistent estimate of standard deviation of the loss differential.[13] The only problem may be a serial correlation of forecast errors causing the loss differential to be serially correlated as well. For this reason, a robust estimation of variance should be employed; we used an estimate of the asymptotic long-run variance.

[13] When using the loss function in the form $L(e_{1t}) = e_{1t}^2$, $\bar{d}$ is simply the difference $MSE_1 - MSE_2$.

An easy way how to calculate the test statistic is to make a linear regression of the loss differential (on the intercept only) using heteroskedasticity and autocorrelation consistent (HAC) standard errors. It is possible to conduct both one-sided and two-sided tests. Initially, Diebold & Mariano (1995) suggested comparing the test statistic with standard normal distribution.

Harvey et al. (1997) proposed few changes based on a set of tests on medium sized data samples. Firstly, they proposed a modification of DM statistic (MDM) in the following way:

$$\mathrm{MDM} = \sqrt{\frac{n+1-2h+\frac{h(h-1)}{n}}{n}} \times \mathrm{DM} \tag{4.27}$$

where $h$ is the forecast horizon and $n$ the number of forecasts. They claim that by this modification, the estimator of variance of $d$ will be unbiased of order $n^{-1}$. For one-step-ahead forecasts, the pre-multiplying coefficient collapses to:

$$\sqrt{\frac{n+1-2h+\frac{h(h-1)}{n}}{n}} \overset{h=1}{=} \sqrt{\frac{n-1}{n}} \tag{4.28}$$

so the final version of MDM statistic for one-step-ahead forecasts is:

$$\mathrm{MDM} = \sqrt{\frac{n-1}{n}} \times \mathrm{DM} \tag{4.29}$$

Also, in analogy with standard tests based on a sample mean, the second change Harvey et al. (1997) proposed was to compare the values of test statistic with critical values of Student $t_{n-1}$ distribution rather than standard normal distribution. Based on simulations on samples of moderate sizes, they showed that the Modified Diebold Mariano test has better size properties than the original one.

#### 4.3.3.2 Clark-West test

The second test of equal forecast accuracy is based on Clark & West (2007). Their test is designed for the case when one of the compared models is a benchmark parsimonious model that is compared with bigger model that nests the first one. Under the null hypothesis, the parsimonious model generates the data; for this reason, the bigger model introduced a noise into its forecasts because it estimates parameters that are equal to zero in the whole population. The authors proposed a method how to clean the MSE of the bigger model from the noise to allow a comparison.

Similarly to the Diebold-Mariano test, a loss differential series is created as a difference of forecast errors of model 1 (benchmark) and model 2 (the bigger one), but an additional adjustment is made to control for the noise[14]:

$$d_t = (y_t - \hat{y}_{1t})^2 - [(y_t - \hat{y}_{2t})^2 - (\hat{y}_{1t} - \hat{y}_{2t})^2] \tag{4.30}$$

where in fact $(y_t - \hat{y}_{1t}) = e_{1t}$ and $(y_t - \hat{y}_{2t}) = e_{2t}$; $(\hat{y}_{1t} - \hat{y}_{2t})^2$ is the adjustment term. The null hypothesis is the equality of MSE of both models with an alternative hypothesis that the bigger model (model 2) has a lower MSE than model 1. The innovation of Clark & West (2007) compared with the Debiold-Mariano test is to take into account the adjusted difference of MSE.[15] Thy null hypothesis should be rejected if the adjusted difference is significantly positive.

When used in practice, they proposed to regress the series $d_t$ on a constant and use t-statistic for the intercept (with the null hypothesis that the constant is equal to zero). Therefore, the Clark-West test statistic (CW) is of a form:

$$\mathrm{CW} = \frac{\bar{d}}{\hat{\sigma}_d} \tag{4.31}$$

The null hypothesis is rejected if the test-statistic is greater than 1.282 (one-sided test at a 10% significance level) or 1.645 (one-sided test on a 5% significance level) if the standard normal distribution is applied. In our analysis, though, we use critical values of the Student $t_{n-1}$ distribution to take the sample size into consideration. As Clark & West (2007) say, usual least squares standard errors can be used when one-step-ahead forecasts are compared. If the forecast errors are autocorrelated, consistent standard errors should be used. Clark & McCracken (2011) also generally recommend using HAC standard errors when comparing more-step-ahead forecasts.

[14] We slightly changed the original notation of Clark & West (2007) to allow for a direct comparison with the procedure of Diebold-Mariano test.

[15] The sample mean of $d_t$ for CW test: $\bar{d} = \mathrm{MSE}_1 - (\mathrm{MSE}_2 - \mathrm{adj})$, where $\mathrm{adj} = \frac{1}{n}\sum_{t=1}^{n}(\hat{y}_{1t} - \hat{y}_{2t})^2$.

## 4.4 Logit model

The aforementioned methods of time series modeling (VAR, ARIMA) are used to capture quantitative characteristics of the dependent variable. Sometimes, it is useful to analyze the time series from a qualitative point of view, such as if the value of the time series is above or below its long-time average.

In the context of our study, this will be used to model the state of the economy (economic downturns), which means converting a time series into a binary form, and getting a dichotomous dependent variable. To model dichotomous dependent variables, probit and logit models are usually used. The paragraphs below provide a description of logit model based on Maddala (2001). The essence of a logit model is the idea that even though we observe variable $y$ that takes only two values:

$$y = \begin{cases} 1 \\ 0 \end{cases} \tag{4.32}$$

there exists a latent variable $y^*$ that determines the value of $y$:

$$y = \begin{cases} 1 \cdots & \text{if } y^* > 0 \\ 0 \cdots & \text{otherwise} \end{cases} \tag{4.33}$$

This latent variable $y^*$ is driven by a formula:

$$y_t^* = \beta_0 + \sum_{j=1}^{k} \beta_j x_{tj} + u_t = X_t\beta + u_t \tag{4.34}$$

where $\beta_i$ are coefficients and $x_i$ are explanatory variables, denoted as $X$ and $\beta$ in the matrix form. Probit model assumes the error term $u_t$ to be normally distributed, logit model assumes the error term to have a logistic distribution.

From expressions (4.33) and (4.34), it follows that:

$$P_t = \text{Prob}(y_t = 1) = \text{Prob}(u_t > -X_t\beta) = 1 - F(-X_t\beta) = F(X_t\beta) \tag{4.35}$$

where $F$ is a cumulative distribution function of $u$. The last identity follows from the symmetry of this distribution. Observed $y_t$ is thus a realization of a binomial process with a probability given by expression (4.35). In the case of logistic distribution of the error term, the probability is:

$$F(X_t\beta) = \frac{\exp(X_t\beta)}{1 + \exp(X_t\beta)} \tag{4.36}$$

which can be rearranged:

$$\ln \frac{F(X_t\beta)}{1 - F(X_t\beta)} = X_t\beta \tag{4.37}$$

and finally:

$$\ln \frac{P_t}{1-P_t} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{tj} \tag{4.38}$$

Left side of this equation is called log-odds ratio, and it is a linear function of explanatory variables.

The model is estimated using maximization of a likelihood function. Because normal and logistic distribution are very similar, only the logistic one has heavier tails, the results of both models will be very similar unless large samples of data are available, meaning a larger accumulation of extreme values (Maddala, 2001).

When evaluating the fit of the model, using classic $R^2$ is problematic since the predicted values $\hat{y}$ are probabilities and the real values $y$ are either 0s or 1s. For this reason, so called pseudo-$R^2$s are used in the case of logit models. They are based on likelihood ratios, comparing the maximum of likelihood function of a model with and without restrictions $L_R$ and $L_{UR}$; the restriction is that all coefficients $\beta_i$ are equal to zero ($\beta_i = 0, i = 1, \ldots, k$).

One of the most commonly used is McFadden $R^2$:

$$\text{McFadden } R^2 = 1 - \frac{\ln L_{UR}}{\ln L_R} \tag{4.39}$$

When explaining dichotomous dependent variable, a natural candidate to assess the fit a model is to measure the number of correctly predicted cases. As mentioned above, predicted values $\hat{y}$ are probabilities that $y = 1$. A value $\hat{y} > 0.5$ can be understood as a prediction that $y = 1$ and $y = 0$ otherwise. Then, it is possible to compare competing models using the share of correct predictions:

$$\begin{gathered}\text{\% of correct predictions} = \frac{\text{correct predictions}}{\text{total number of observations}} \\ \text{\% of correct downturn predictions} = \frac{\text{correct downturn predictions}}{\text{total number of downturns}}\end{gathered} \tag{4.40}$$

## 4.5 Seasonality

Most of economic time series – both financial and macroeconomic – contain some kind of seasonality; and this is true also for time series of Google data. Seasonality is a component of time series with a repetitive character with a particular frequency – quarterly, monthly, or even daily. If the data contain seasonality and this fact is ignored when making models, it may result in misspecifications. Seasonally adjusting the data can thus help in uncovering true relations in the data.

In practice, two methods for seasonal adjustments are used and recommended across statistical offices of the European Union: X-12-ARIMA (developed by the United States Census Bureau) and TRAMO/SEATS (developed by V. Gómez and A. Maravall of the Bank of Spain). TRAMO/SEATS is used for example by the Czech Statistical Office, and we applied it in our analysis as well. It is freely available as an add-in for Gretl software.[16] Gretl is open-source statistical software used for all model estimations in our thesis.[17]

TRAMO stands for "Time Series Regression with ARIMA Noise, Missing Observations and Outliers", SEATS stands for "Signal Extraction in ARIMA Time Series". As Gómez & Maravall (1997) describe, TRAMO prepares the data for seasonal adjustments (e.g. detects and corrects for outliers or missing observations) and SEATS estimates components of the time series (trend, cycle, seasonal, irregular) using signal extraction techniques applied to ARIMA model.

Even though most of the time series downloaded from official sources were already seasonally adjusted, as Enders (2003) points out, it does not mean that there is no seasonality in the analyzed data sample. The original adjustment was probably made over a different time span and also could be done by a different method. For this reason, we seasonally adjusted also time series that were acquired as adjusted.

[16] http://gretl.sourceforge.net/tramo/tramo-seats.html
[17] http://gretl.sourceforge.net/ ; in addition, forecast evaluation was made in Microsoft Excel

# 5 Unemployment

Unemployment is an area of research where Google Econometrics has quickly found its use. The reasons for this are obvious: usually, the official rate of unemployment is published with a significant delay after the end of the reference month. It is for example three months in Italy, two and half months in Turkey or two months in Israel and the United States. In other countries, such as Germany, France, or even the Czech Republic (as mentioned in the data description), this delay – one month – is not that severe, but still significant, and early precise predictions of the unemployment rate are useful.

At the same time, using the internet as a means of searching for job is common among unemployed internet users; it was for example at least 60% during the analyzed period in the Czech Republic and even more in other compared countries.[18] At the same time, the use of a search engine to find information is a common skill among internet users, so it is reasonable to assume than at least some of the internet job search is conducted using Google search engine.

If people indeed do this, using appropriate search queries will help in assessing the total rate of unemployment. The goal of this chapter is to test the following hypothesis:

> Google search query data can be used to estimate the current Czech unemployment rate in advance compared with other methods.

In the related studies for other countries, the authors differed in the choice of search queries to analyze. Some researchers used one search query only, some used more queries or a whole category. For one query, one of the most popular is "job" or "jobs". General motivation is the widespread use of such word among people looking for job and also higher search popularity compared with alternatives. For example D'Amuri & Marcucci (2010) used a query "jobs" for the United States, Fondeur & Karamé (2013) used "emploi" (meaning "job") for France. D'Amuri & Marcucci (2012) also tested entries "collect unemployment" and "job center" for the United States and D'Amuri (2009) used "job offers" for Italy.

[18] Charts are provided in Appendix A in the description of internet skills and activities.

Instead of just one query, Suhoy (2009) used a whole category called "Human resources" for Israeli data; other authors rather used predetermined set of individual queries. Askitas & Zimmermann (2009) based their analysis on four queries with the aim to cover several groups of people looking for a job in Germany: (a) "unemployment office" to cover those who lost their job; (b) "unemployment rate" for those generally concerned with the topic; (c) "personal consultant" for qualified people in danger of losing their job due to restructuring; and (d) names of popular websites with job offers to cover those actively looking for a job.

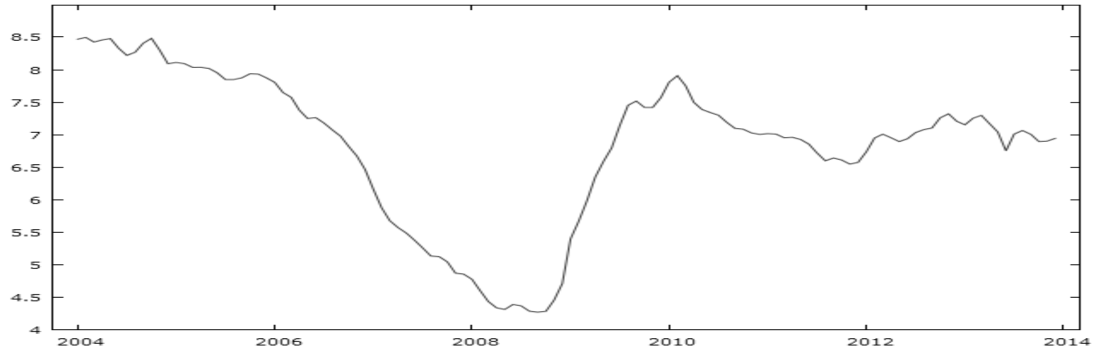
Chadwick & Sengul (2012) also made own list of queries related to unemployment for Turkish data: "job search", "job offers", "CV", "career", "unemployment", "unemployment insurance" and also names of websites with job offers. They aggregated information from all the data using principal component analysis and used first few components; Askitas & Zimmermann (2009) used all of their queries individually.

The goal of our analysis is to assess the predictive power of Google data for the Czech unemployment rate. For this, we will employ the following empirical strategy: firstly, we find the most appropriate ARIMA model for the unemployment rate using Box-Jenkins methodology. Next, we use Google data and other control variables as additional explanatory variables in the chosen ARIMA model. We compare the explanatory power of individual variables based on the quality of out-of-sample forecasts, similarly to the related literature. AR model was used for example by Choi & Varian (2012) and Chadwick & Sengul (2012), ARIMA by D'Amuri & Marcucci (2012) and Suhoy (2009).

## 5.1 Unemployment rate

Czech unemployment rate is published by the Czech Statistical Office (CZSO), the data are based on a quarterly Labor Force Survey. Even though detailed data are available only with a quarterly frequency, basic information is published monthly with a delay of one month; we use this monthly unemployment rate for the age category 15–64 over the period from January 2004 to December 2013. The time series is depicted in Figure (5.1). Even though the original time series was acquired as seasonally adjusted, we used TRAMO/SEATS to remove any remaining seasonality on a given sample. The same was done also for all other variables.

**Figure 5.1: Seasonally adjusted Czech unemployment rate (2004–2013).**




Source: CZSO

After a gradual decline in the rate of unemployment to 4.5% in the middle of 2008, it quickly rose close to its initial values after the economic downturn during 2009. We will model this time series in accordance with the Box-Jenkins methodology; the first step is to achieve stationarity of the time series. Based on visual analysis, the series does not seem to be stationary, even though this is probably caused by the break described above. To formally test stationarity of the series, Augmented Dickey-Fuller (ADF) and KPSS tests were conducted; the results are presented in Table (5.1).

**Table 5.1: Stationarity tests for the Czech unemployment rate**

| | **Test statistic** | **P-value** |
|---|---|---|
| **ADF** | -0.7049 | 0.4117 |
| **KPSS** | 0.1867 | > 0.1 |

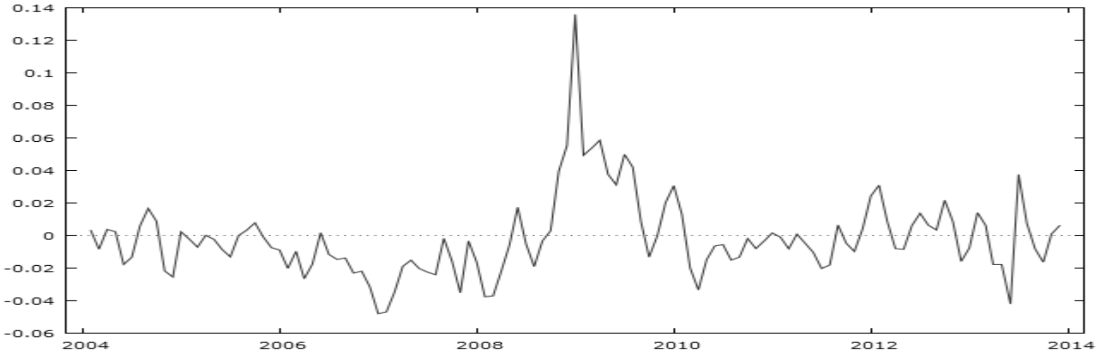
Source: author's calculations

The null hypothesis of ADF test is the presence of unit root in the series, meaning non-stationarity. In this case, we cannot reject the null hypothesis. On the other hand, the null hypothesis of KPSS is stationarity of the series, and we cannot reject it either. Put together with the visual analysis, and also with the fact that some other series of additional explanatory variables were found to be non-stationary, we decided to transform the data and conduct the analysis on logarithmic differences of unemployment rate:

$$ld\_y_t = \ln(y_t) - \ln(y_{t-1}) = \ln\left(\frac{y_t}{y_{t-1}}\right) \qquad (5.1)$$

The time series of log-differences is shown in the Figure (5.2) and the results of stationarity tests are displayed in Table (5.2). The transformed series appears more stationary, even though the break in years 2009 is preserved. In this case, however, we reject the null hypothesis of the presence of unit root for the ADF test on 5% significance level; for the KPSS test, the value of test statistic further decrease, so we cannot reject the null hypothesis of stationarity. Because we achieved stationarity by log-differencing the data, the same transformation is applied to all other series. One disadvantage is the loss of the first observation.

**Figure 5.2: Logarithmic differences of the Czech unemployment rate (2004–2013)**




Source: CZSO, author's calculations

**Table 5.2: Stationarity tests for the log-differences of unemployment rate**

| | **Test statistic** | **P-value** |
|---|---|---|
| **ADF** | -2.4956 | 0.0122 |
| **KPSS** | 0.1368 | > 0.1 |

Source: author's calculations

The second step is the identification of the model's specification using information criteria. When looking at the whole data sample, Akaike information criterion (AIC) suggests to use a large ARIMA(3,2) model; Schwartz (BIC) and Hannan-Quinn (HQIC) recommend parsimonious AR(1) model specification. Firstly, we check the parsimonious version; this choice was confirmed also by the Ljung-Box test of remaining linear dependencies in residuals of AR(1) model. For the maximum number of 6 lags, the p-value for the null hypothesis of no remaining linear dependencies is 0.2887 – we cannot reject the null (the same applies for other maximum numbers of lags).

The choice of parsimonious model is further confirmed by looking at the information criteria on rolling window subsamples of a length of 60 observations (relevant for further analysis). The AR(1) model was chosen in the vast majority of cases by all three information criteria (smaller deviations appeared only by the end of the sample). Table (5.3) shows the Gretl estimates of the AR(1) model for log-differences of the unemployment rate. The estimate of coefficient of the first lag is 0.68, which is in agreement with previously rejected hypothesis about the presence of unit root. However, the assumption about normality of residuals was rejected. The resulting model is AR(1) without constant for log-differences, which means ARIMA(1,1,0) without constant for the levels of unemployment rate.

**Table 5.3: Estimate of AR(1) model for log-differences of unemployment rate**

```
Model 1: ARMA, using observations 2004:02-2013:12 (T = 119)
Estimated using Kalman filter (exact ML)
Dependent variable: ld_U_sa
Standard errors based on Hessian

             coefficient   std. error     z      p-value
  -------------------------------------------------------
  phi_1       0.679484     0.0660836    10.28   8.48e-025 ***

Mean dependent var  -0.001667   S.D. dependent var   0.024521
Mean of innovations -0.000499   S.D. of innovations  0.017834
Log-likelihood       310.0104   Akaike criterion    -616.0208
Schwarz criterion   -610.4625   Hannan-Quinn        -613.7638

                        Real  Imaginary    Modulus  Frequency
  -----------------------------------------------------------
  AR
    Root  1           1.4717     0.0000     1.4717     0.0000
  -----------------------------------------------------------
```

Source: author's calculations in Gretl

## 5.2 Additional explanatory variables

As early indicators, variables that are collected with a higher frequency or that are available in advance are used – for example it is the data about initial jobless claims in the United States, published weekly and measuring the number of people filing for unemployment benefits (D'Amuri & Maruccuci, 2010). An analogy for the Czech Republic is the share of unemployed which is based on the number of people registered in the Labour office. These values are available several weeks before the official unemployment rate and may indicate its future development.

Unemployment is also closely related to the situation of the whole economy and its cycle. For this reason, indicators either measuring some part of the economy or created to anticipate changes in the economic situation will be used as control variables. Index of industrial production is usually used to assess the economic situation (for example OECD used it as a reference series for the creation of their CLIs in the past); this index can represent the macroeconomic situation. Confidence indicators capture expectations about the future state of the economy by people in business and by consumers. Even more directly, one of the questions consumers are asked is the expected total unemployment. CLIs is also designed to capture the changes in the economy in advance.

The following variables will be used as controls during the analysis:

- Share of unemployed

- Index of Industrial Production
- Confidence Indicators
- Composite Leading Indicators

### 5.2.1 Google data

Because categories are not available for the Czech Republic, we used individual search queries of Google data. Similarly to the related literature, we analyzed queries that are likely to be used by a person looking for a job on the internet or an unemployed. Generally when choosing such query, the accent was put on simple ones, but in some cases, we used a combination if the simple one did not provide a complete series. The paragraphs below provide description of the data and selection motivation.

• "**Job**" ("práce" in Czech; also means "work", "labour", but even "thesis")
Similarly to the reviewed literature, we expect this query to be of a widespread use among people looking for a job on the internet. Its popularity is confirmed also by comparison with other possible queries reviewed below. Google data of this entry give information about all queries that contain this (Czech) word, some are connected to job search and we will use them as well (e.g. "job offers" or "Labour office").

There are also queries unrelated to job search, such as words containing word "thesis" ("bachelor thesis", "diploma thesis", "doctoral thesis", etc.). We conducted the analyses also on the data cleaned from this effect of unrelated searches, but we present only the results for the basic query for the following reasons: (a) the results did not differ significantly; (b) cleaning the data from such noise would never be complete as there is too many possible noise terms – attempting to clean for their effects would make the use of Google data in practice too complicated.

Another concern with Czech data is declension in the Czech language – not only plural and singular form, but also changes of the word depending on the case. We used singular form for most queries and the entries are in the nominative case. Lastly, the concern is the use of diacritic marks (accute accent, caron) among Czech internet users. Preliminary analysis suggests that some queries are indeed often searched for even without diacritic marks, so the analysis was conducted on time series with such version of the words. Similarly to the case of cleaning the data, the results did not differ much and it would complicate the use of Google data in practice, these results are not presented.

• "**Employment**" ("zaměstnání" in Czech, also translated as "work" or "job")
This word (in Czech) has a similar meaning to "job" described above, even though it is searched for to a lesser extent. This is obvious from the Figure (5.3) that shows big variance of the time series in the beginning of the sample. However, this word is likely to contain less noise than "job" since its meaning is strictly related to a working practice.

• **"Job offers"**
The exact string analyzed in Czech was "volné místo" + "volná místa" + "pracovní místo" + "pracovní místa" + "pracovní pozice" + "nabídka práce" + "nabídky práce". Meaning of these entries is "job offer", "vacancy", "job position" in their singular and plural form. This rather extensive list of entries had to be used to obtain more complete series of Google data. As Figure (5.3) depicts, the series is still very volatile in the beginning of the sample even after the inclusion of more terms.

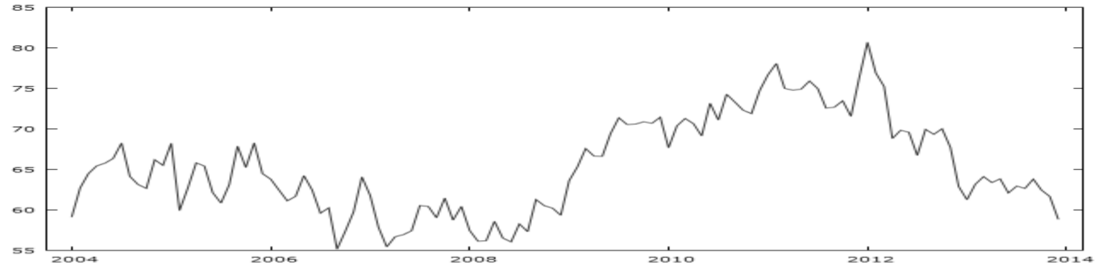
• **"Labour office"** (in Czech, the string was "úřad práce" + "pracovní úřad" + "úp")
The abbreviation of the name of the office was included as well as a slightly incorrect, but often used name for this institution dealing with unemployed people. An analysis was conducted also using the raw version with slightly worse results. This search query should represent people who are unemployed, look for a job, and either ask for social benefits or ask the state to cover their social and health insurance. Labour Office is the institution keeping track of registered unemployment.

• "**CV**" ("životopis" in Czech, also means "biography")
This query was also chosen to represent people actively searching for a job, since a CV is one of the requirements usually asked by any potential employer, and people look for advice and tips on what should be included in the CV and how to write it. Time series for this query is also shown in Figure (5.3).

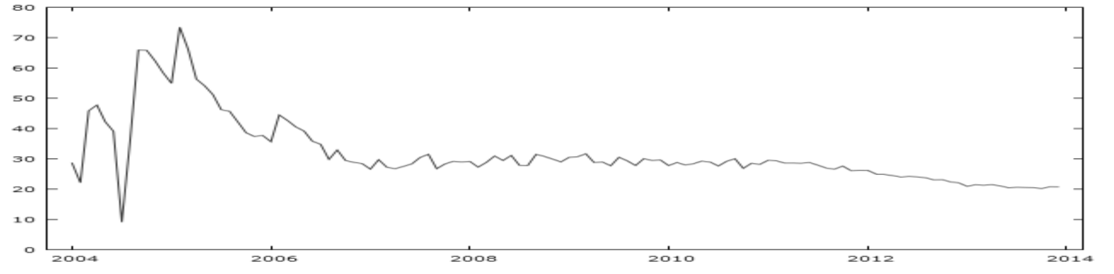
### 5.2.2 Preliminary analysis

We use cross-correlogram for logarithmic differences to see the short term relationship between the unemployment rate and analyzed Google search queries. It is an analogy to Autocorrelation function which shows the interconnection between the current value of one variable and the value of the same

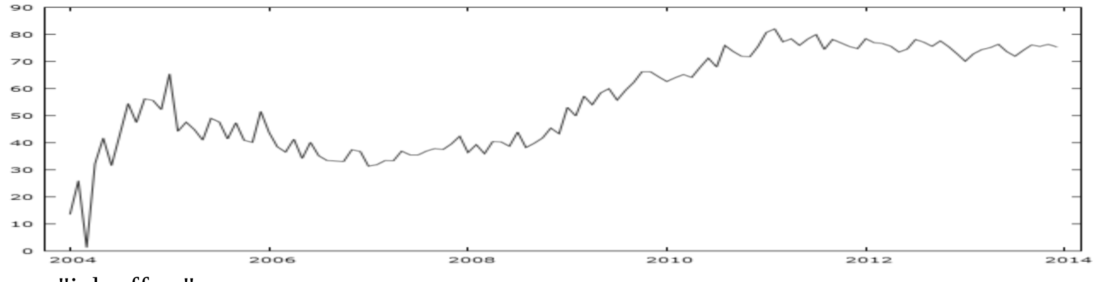
**Figure 5.3: Seasonally adjusted search volume indices of analyzed Google queries**



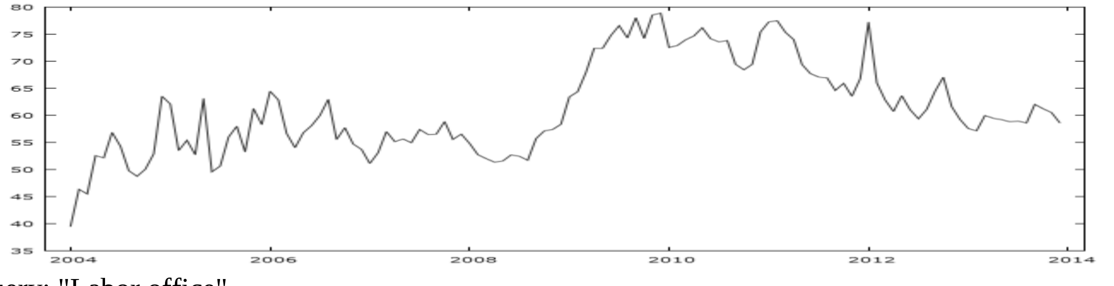

Query: "job"



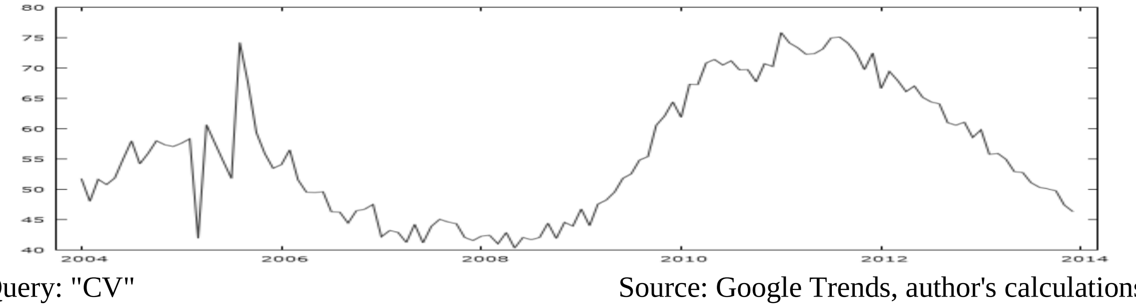

Query: "employment"




Query: "job offers"




Query: "Labor office"



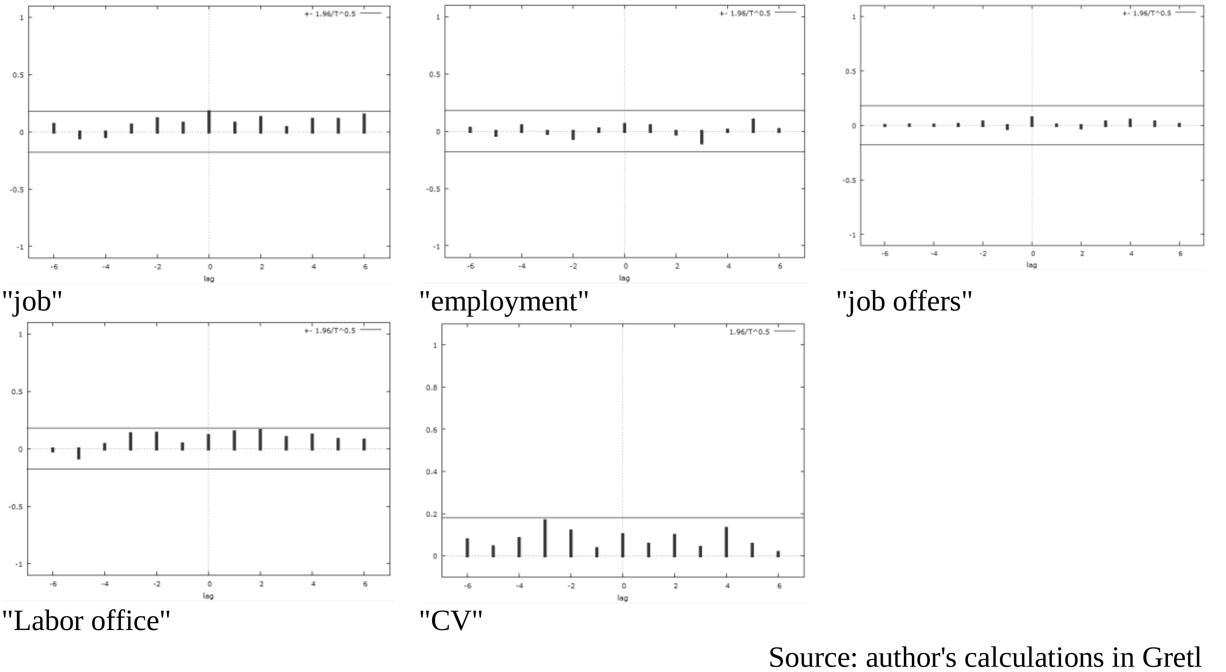

Query: "CV"

Source: Google Trends, author's calculations

variable $k$ periods behind. Cross-correlogram shows the same, except it is the interconnection between the current value of one variable and lags of the other variable.

Figure (5.4) shows cross-correlograms for logarithmic differences of Google search query data vis-à-vis the logarithmic differences of the unemployment rate. For all queries and most of the lags, the correlation coefficients are positive. The strongest relationship – even though still non-significant or only slightly significant – is present for ”job” and ”Labour office”; stronger relationship prevails on the right hand side of the chart, meaning that changes in the search for these queries lead changes in unemployment. The opposite is true for ”CV” and the correlation is almost zero for other two queries.

**Figure 5.4: Cross-correlograms between log-differences of unemployment and log-differences of individual queries (2004–2013)**

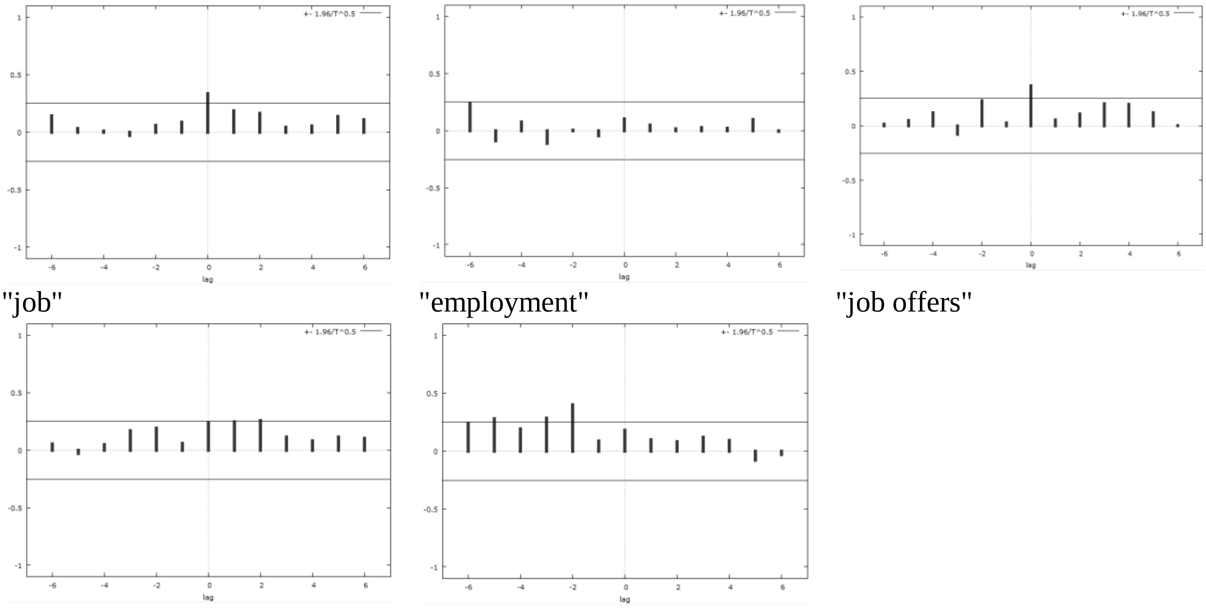



Source: author's calculations in Gretl

Explanation: Correlations for lagged values of Google data (to the current unemployment rate) are indicated in the right hand side of each chart, correlations for lagged values of unemployment rate (to the current value of Google data) are shown on the left hand side of each chart.

But importantly, when looking only at the second half of the analyzed period, these relationships changed. Figure (5.5) shows this for the period 2009–2013 – most notably, the correlation increased strongly for ”job” and ”Labour office” and also for ”job offers”. For each of them, the contemporaneous correlation became significant; the same is true also for the first two lags of ”Labour office”. On the other hand, the relationship of ”Employment” did not improve and the reverse causal connection for ”CV” became more apparent in the second half of the analyzed period.

**Figure 5.5: Cross-correlograms between log-differences of unemployment and log-differences of individual queries (2009–2013)**



"job" "employment" "job offers"

"Labor office" "CV"

Source: author's calculations in Gretl

Explanation: Correlations for lagged values of Google data (to the current unemployment rate) are indicated in the right hand side of each chart, correlations for lagged values of unemployment rate (to the current value of Google data) are shown on the left hand side of each chart.

This shift is probably caused by several factors. As charts in Figure (5.3) depicted for most of the queries, their time series were noisy by the beginning of the analyzed period, meaning Google data did not contain strong signal; this improved during time. Also, during the whole period, the internet has become more popular when searching for job in the Czech Republic, and Google increased its share on the Czech search engine market. All in all, this might lead to a better explanatory power of Google data.

## 5.3 Nowcasting

The preliminary analysis has shown promising results for some of the queries, especially during the second half of the analyzed period. To see the real explanatory power, we test the quality of out-of-sample predictions using these data compared with other control variables. The main advantage of Google data – when forecasting or when assessing the current state of the economy – is that they are published with practically no delay. The data are updated weekly and can be downloaded any time during the month.

This is one of the main differences compared with most of the control variables and the reason why Google Econometrics has gained its popularity in the field of short term predictions or nowcasting. Table (5.4) presents the availability of variables used in our analysis during the first quarter of 2014 with marked publication dates. In other months, the day of the week and other factors may cause the publication day to differ. Since we calculate monthly values of Google data as an average over a given month, we denoted the last day of every month as its "publication" date.

**Table 5.4: Diagram of variables' publication dates (during the 1st quarter of 2014)**

| | January 2014 | | February 2014 | March 2014 | April |
|---|---|---|---|---|---|
| Unemployment Rate (CZSO) | Nov. | Dec. | Jan. | | Feb. |
| Confidence Indicators (CZSO) | | Jan. | Feb. | Mar. | |
| Search query data (Google) | | Jan. | Feb. | Mar. | |
| Share of Unemployed (LO) | Dec. | | Jan. | Feb. | |
| Index of Industrial Production (CZSO) | Nov. | | Dec. | Jan. | |
| Composite Leading Indicators (CLIs) | Nov. | | Dec. | Jan. | |

Source: CZSO, OECD, Labor Office

Our goal is to model the unemployment rate with an AR(1) process, the explanatory variable is the first lag of the dependent variable. Therefore, for one-step-ahead predictions, we need to know this lag, and as indicated in the table, the unemployment rate is published by the end of next month (only November data are published with a longer delay). By the end of February, the unemployment rate for January is published and it can be used to forecast – or rather nowcast – the February unemployment rate.

When augmenting the AR(1) model with additional explanatory variables to improve nowcasting performance, we need to take into account their availability. By the end of February, the following values are known: February (lag 0) – Confidence Indicators, Google data; January (lag 1) – Share of unemployed; December (lag 2) – Index of Industrial Production, Composite Leading Indicators. For longer forecast horizon, the available lag would increase accordingly.

### 5.3.1 Empirical Strategy

Firstly, we divide the data into an in-sample and out-of-sample parts of lengths $R$ and $P$. The first part of the sample will be used for initial estimates of the models that will be used for forecasting; the second part of the sample will be used to compare forecasting qualities of different models. In accordance with recommendations described in the methodological section, we divided the sample in half, having $R = P = 60$ for the following reasons: Clark & McCracken (2011) recommend having the ratio of $P/R$ at least 1 or even higher, similarly to Hansen and Timmermann (2012). And Enders (2003) recommends the number of observations used for forecasting with ARMA models to be at least 50.

On the initial sample of length $R = 60$ ending with December 2008, AR(1) model is estimated:

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t \tag{5.2}$$

Based on the estimates of parameters of this model, a prediction is made for January 2009:

$$\hat{y}_{JAN} = \hat{\alpha}_1 y_{DEC} \tag{5.3}$$

because the expected value of $\hat{\varepsilon}_{JAN} = 0$. When using additional explanatory variables, the model is of a form:

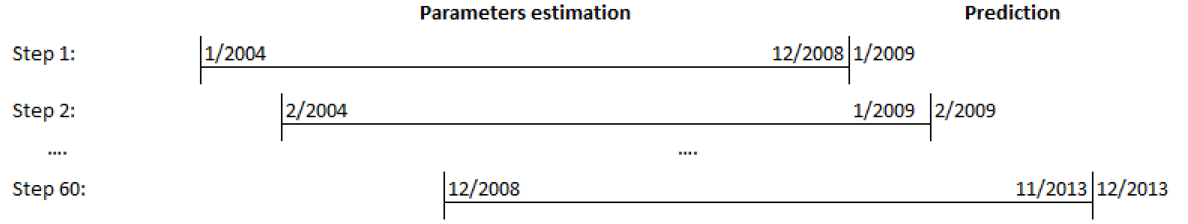
$$y_t = \alpha_1 y_{t-1} + \gamma_1 X_{t-l} + \varepsilon_t \tag{5.4}$$

where $X_{t-l}$ is the additional explanatory variable known at time $t$, $l$ denotes the lag of such variable. More explanatory variables can be included. Assuming $l = 0$, the nowcasting relationship is:

$$\hat{y}_{JAN} = \hat{\alpha}_1 y_{DEC} + \hat{\gamma}_1 X_{JAN} \tag{5.5}$$

This procedure is repeated for forecasts until the end of the out-of-sample period $P$ times. The rolling window scheme is used, meaning that the initial window of length $R = 60$ moves gradually with each step. For example West (2006) recommends this scheme if we want to defend against parameter drift which is difficult to model directly. As we saw in the preliminary analysis, there was indeed a change in relations between Google data and the unemployment rate. The procedure is graphically described in the Table (5.5).

**Table 5.5: Diagram of rolling window scheme for out-of-sample forecasting**




Source: author's description

In reality, the number of observations used for model estimates will not always be $60$. As mentioned above, we lost one observations due to differencing the data, also the time series for the Share of unemployed is available only since 2005 (it is shorter by 12 observations). In addition, when using lags of the presented variables, additional loss of observations was introduced. Therefore, the models will be estimated such that the maximum number of observations that belong to the particular window will be used. Nevertheless, the actual number of observations used for estimates will converge quickly to 60.

Using this procedure, 60 out of sample forecasts are calculated for each of the models. In accordance with the description provided in the methodological section, the Mean Squared Error (MSE) is calculated for each of the models and the Clark-West test of equal predictive accuracy for nested models is used to assess the significance of MSE differences.

### 5.3.2 Results

For Google data, we allowed the maximum lag to be 6 months to test the theoretical interconnection between the unemployment rate and contemporaneous activities of internet users. For other explanatory variables, the lag up to 11 months was allowed, because some of the indicators are designed to anticipate the real economic situation in advance; this was confirmed by results. Also, we took into account the theoretical relationship between a given variable and the unemployment rate (for example that an increase in confidence should not predict rising unemployment rate) and the availability of the data at the time of nowcasting as shown in Table (5.4).

Results are presented for the best performing lag for each variable. Firstly, the MSE of the baseline autoregressive model AR(1) is presented. Then, for each variable, MSE of an ARX model (AR with an additional explanatory variable) was calculated for the denoted lag. A relative MSE was calculated:

$$\%\ \text{MSE of AR1} = \frac{\text{MSE}_{\text{ARX}}}{\text{MSE}_{\text{AR(1)}}} \tag{5.6}$$

And finally, the potential improvement of the augmented model was formally assessed using the Clark-West test; the null hypothesis is their equal predictive accuracy with the alternative that the bigger model makes better predictions. The test statistic, p-value and its significance are shown in Table (5.6). A general observation is that additional explanatory variables did not improve much the baseline AR(1) model in terms of predictive accuracy. The best model has MSE only by 7% smaller than the benchmark, and two models with 3% improvements follow. On the other hand, most of these improvements are statistically significant at 10% or even 5% confidence level.

**Table 5.6: Out-of-sample nowcasting of unemployment (period 2009–2013)**

| Out-of-sample period | | 2009 - 2013 |
|---|---|---|
| Autoregressive model AR1 (benchmark) | MSE | 0.0005237 |
| Additional explanatory variable | Lag | % MSE of AR1 |
| Consumer Confidence Indicator | 7 | **97.0%**<br>2.30<br>(0.013)<br>** |
| Composite Confidence Indicator | 7 | **98.8%**<br>1.06<br>(0.147) |
| Index of Industrial Production | 8 | **93.1%**<br>2.30<br>(0.013)<br>** |
| Composite Leading Indicators | 8 | **99.0%**<br>1.52<br>(0.067)<br>* |
| Share of Unemployed | 2 | **99.4%**<br>1.84<br>(0.035)<br>** |
| Google query: "job" | 0 | **97.9%**<br>1.75<br>(0.043)<br>** |
| Google query: "employment" | 0 | **99.0%**<br>2.17<br>(0.017)<br>** |
| Google query: "job offers" | 0 | **98.8%**<br>1.19<br>(0.119) |
| Google query: "Labour office" | 2 | **99.1%**<br>1.31<br>(0.098)<br>* |
| Google query: "CV" | 4 | **96.7%**<br>1.82<br>(0.037)<br>** |

Source: author's calculations

Explanation: The percentage value shows the relative MSE of the augmented ARX model to the benchmark AR(1) model, value lower than 100% implies improved forecasts when additional variable is introduced. Next rows show the Clark-West test statistics, p-value and significance of this test. The null hypothesis is equal predictive accuracy of both models; we cannot reject it for high p-values, meaning that potential improvement is not statistically significant. Alternative hypothesis is superior forecasting quality of the larger model. Significance is denoted by asterisks (* for 10%, ** for 5%, and *** for 1% significance level).

Concerning control variables, the Index of Industrial Production performed the best, followed by Consumer Confidence Indicator. Composite Leading Indicators and the Share of unemployed improved the baseline model only slightly and the improvement by Composite Confidence Indicator was not statistically significant. It is good to notice that from control variables, the best performance was displayed by 7$^{th}$ (confidence) or 8$^{th}$ lag (industrial production, CLIs). There seems to be a consistent relationship between these variables – designed to capture the state of the whole economy – and the unemployment rate.

Compared with control variables, Google data performed relatively well. Even though the improvement was only mild (maximum of 3% for "CV" and 2% for "job"), it was statistically significant in all cases except for the "job offers" query (p-value was very close to 0.1). In the case of search query data, contemporaneous values or past few lags performed the best, which is consistent both with theoretical expectations and with cross-correlogram analysis.

Visual observations of cross-correlograms also suggested that the interconnection between Google data and unemployment improved during the course of the analyzed period. Also, some of the Google time series showed noise and volatility at their beginning. To see if the explanatory power of Google data (or control variables) changed within the out-of-sample period, we examined its subsamples. Table (5.7) shows the results analogous to the Table (5.6), only various out-of-sample periods are considered. The first column shows results identical to Table (5.6), the second column considers an out-of-sample period of 2010–2013, and the third of 2011–2013. At the same time, the window for estimating coefficients is always $R = 60$, so for the out-of-sample 2010–2013, the initial estimate was made over 2005–2009 window, and so on.

The results indicate that indeed the improvement in MSE relatively to the benchmark AR(1) model was better when looking at out-of-sample of 2010–2013 and even more for 2011–2013. This means that for initial estimations, observations from the first year (2004) or first two years (2004–2005) were not used – implying that the noise contained in Google data in its early observation probably played a significant role. This is also indicated by the fact that the performance of control variables – except for composite leading indicators – did not follow this pattern of improvement.

For example the query "job" belongs among the best when looking at subsamples, similarly "job offers" improved the benchmark forecasts significantly. The improvement by "employment" was still mild but significant at 5% level, and improvements made by "Labour Office" and "CV" were only slightly insignificant. From the control variables, Index of Industrial Production performed the best this time together with Composite Leading Indicators.

**Table 5.7: Out-of-sample nowcasting of unemployment (various periods)**

| Out-of-sample period | | **2009 - 2013** | **2010 - 2013** | **2011 - 2013** |
|---|---|---|---|---|
| Autoregressive model AR1 (benchmark) | MSE | 0.0005237 | 0.0002571 | 0.0002887 |
| Additional explanatory variable | Lag | % MSE of AR1 | % MSE of AR1 | % MSE of AR1 |
| Consumer Confidence Indicator | 7 | **97.0%**<br>2.30<br>(0.013)<br>** | **97.8%**<br>1.67<br>(0.051)<br>* | **99.1%**<br>1.23<br>(0.113) |
| Composite Confidence Indicator | 7 | **98.8%**<br>1.06<br>(0.147) | **101.5%**<br>0.43<br>(0.335) | **99.7%**<br>0.85<br>(0.199) |
| Index of Industrial Production | 8 | **93.1%**<br>2.30<br>(0.013)<br>** | **96.3%**<br>1.25<br>(0.109) | **95.4%**<br>1.22<br>(0.115) |
| Composite Leading Indicators | 8 | **99.0%**<br>1.52<br>(0.067)<br>* | **94.0%**<br>1.48<br>(0.073)<br>* | **93.4%**<br>1.52<br>(0.069)<br>* |
| Share of Unemployed | 2 | **99.4%**<br>1.84<br>(0.035)<br>** | **105.7%**<br>1.38<br>(0.087)<br>* | **103.3%**<br>1.95<br>(0.030)<br>** |
| Google query: "job" | 0 | **97.9%**<br>1.75<br>(0.043)<br>** | **95.0%**<br>1.73<br>(0.045)<br>** | **91.3%**<br>2.02<br>(0.025)<br>** |
| Google query: "employment" | 0 | **99.0%**<br>2.17<br>(0.017)<br>** | **98.1%**<br>2.00<br>(0.026)<br>** | **97.8%**<br>1.97<br>(0.029)<br>** |
| Google query: "job offers" | 0 | **98.8%**<br>1.19<br>(0.119) | **95.3%**<br>1.68<br>(0.050)<br>* | **93.6%**<br>1.81<br>(0.039)<br>** |
| Google query: "Labour office" | 2 | **99.1%**<br>1.31<br>(0.098)<br>* | **97.6%**<br>1.35<br>(0.092)<br>* | **97.9%**<br>1.17<br>(0.126) |
| Google query: "CV" | 4 | **96.7%**<br>1.82<br>(0.037)<br>** | **97.0%**<br>1.04<br>(0.153) | **95.9%**<br>1.15<br>(0.130) |

Source: author's calculations

Explanation: The percentage value shows the relative MSE of the augmented ARX model to the benchmark AR(1) model, value lower than 100% implies improved forecasts when additional variable was introduced. Next rows show the Clark-West test statistics, p-value and significance of this test. The null hypothesis is equal predictive accuracy of both models; we cannot reject it for high p-values, meaning that potential improvement is not statistically significant. Alternative hypothesis is superior forecasting quality of the larger model. Significance is denoted by asterisks (* for 10%, ** for 5%, and *** for 1% significance level).

Since some of the Google queries perform the same as the best control variables (or even better), the question is whether it brings some additional information over that already contained in macroeconomic variables. For this reason, an analogous analysis to the previous one was conducted with a different benchmark model. First, the benchmark model was an ARX model with Index of Industrial Production ($8^{th}$ lag); and second, the benchmark model was an ARX model with Composite Leading Indicators ($8^{th}$ lag).

**Table 5.8: Out-of-sample nowcasting of unemployment – combination with the Index of Industrial Production**

| Out-of-sample period | | 2009 - 2013 | 2010 - 2013 | 2011 - 2013 |
|---|---|---|---|---|
| Autoregressive model AR(1) | MSE | 0.0005237 | 0.0002571 | 0.0002887 |
| AR(1) with 8th lag of Index of Industrial Production (ARX) | | 93.1% | 96.3% | 95.4% |
| Additional explanatory variable | Lag | % MSE of ARX | % MSE of ARX | % MSE of ARX |
| Consumer Confidence Indicator | 7 | **96.9%**<br>2.48<br>(0.008)<br>*** | **98.6%**<br>1.60<br>(0.058)<br>* | **100.4%**<br>1.10<br>(0.140) |
| Composite Confidence Indicator | 7 | **100.2%**<br>0.21<br>(0.419) | **101.0%**<br>0.04<br>(0.483) | **99.9%**<br>0.49<br>(0.313) |
| Composite Leading Indicators | 8 | **101.8%**<br>1.32<br>(0.096)<br>* | **95.9%**<br>1.31<br>(0.099)<br>* | **95.5%**<br>1.40<br>(0.086)<br>* |
| Share of Unemployed | 2 | **102.3%**<br>1.67<br>(0.050)<br>* | **106.4%**<br>1.44<br>(0.078)<br>* | **104.6%**<br>1.99<br>(0.027)<br>** |
| Google query: "job" | 0 | **96.9%**<br>2.06<br>(0.022)<br>** | **92.2%**<br>2.11<br>(0.020)<br>** | **89.6%**<br>2.20<br>(0.017)<br>** |
| Google query: "employment" | 0 | **100.0%**<br>0.18<br>(0.427) | **98.7%**<br>1.97<br>(0.027)<br>** | **98.5%**<br>1.87<br>(0.035)<br>** |
| Google query: "job offers" | 0 | **92.4%**<br>2.90<br>(0.003)<br>*** | **92.6%**<br>2.04<br>(0.024)<br>** | **90.4%**<br>2.11<br>(0.021)<br>** |
| Google query: "Labour office" | 2 | **100.4%**<br>0.59<br>(0.278) | **99.6%**<br>0.89<br>(0.189) | **99.9%**<br>0.80<br>(0.214) |
| Google query: "CV" | 4 | **96.6%**<br>1.43<br>(0.079)<br>* | **97.3%**<br>1.00<br>(0.160) | **96.3%**<br>1.08<br>(0.144) |

Source: author's calculations

Explanation: The percentage value shows the relative MSE of the augmented ARX model to the benchmark ARX model, value lower than 100% implies improved forecasts when additional variable was introduced. Next rows show the Clark-West test statistics, p-value and significance of this test. The null hypothesis is equal predictive accuracy of both models; we cannot reject it for high p-values, meaning that potential improvement is not statistically significant. Alternative hypothesis is superior forecasting quality of the larger model. Significance is denoted by asterisks (* for 10%, ** for 5%, and *** for 1% significance level).

The results are shown in Tables (5.8) and (5.9), completely analogous to the Table (5.7), only with a different benchmark model. When looking at control variables, Composite Leading Indicators improves the benchmark ARX with Index of industrial production (Table 5.8) only in the subsamples that do not take into account the first year, which is in accordance with previous results. The only other control variable significantly improving the benchmark in some cases was Consumer Confidence Indicator.

On the other hand, the strength of Google data seems to be in the combination with the control variable. For example queries ”job” and ”job

**Table 5.9: Out-of-sample nowcasting of unemployment – combination with the Composite Leading Indicators**

| Out-of-sample period | | 2009 - 2013 | 2010 - 2013 | 2011 - 2013 |
|---|---|---|---|---|
| Autoregressive model AR(1) | MSE | 0.0005237 | 0.0002571 | 0.0002887 |
| AR(1) with 8th lag of Composite Leading Indicators (ARX) | | 99.0% | 94.0% | 93.4% |
| Additional explanatory variable | Lag | % MSE of ARX | % MSE of ARX | % MSE of ARX |
| Consumer Confidence Indicator | 7 | **100.0%**<br>0.86<br>(0.197) | **97.1%**<br>1.64<br>(0.054)<br>* | **99.6%**<br>1.08<br>(0.144) |
| Composite Confidence Indicator | 7 | **100.5%**<br>0.19<br>(0.426) | **102.1%**<br>-0.08<br>---- | **100.4%**<br>0.26<br>(0.400) |
| Index of Industrial Production | 8 | **95.7%**<br>1.95<br>(0.028)<br>** | **98.3%**<br>0.84<br>(0.201) | **97.5%**<br>0.96<br>(0.171) |
| Share of Unemployed | 2 | **105.6%**<br>1.03<br>(0.153) | **105.3%**<br>1.18<br>(0.123) | **104.0%**<br>1.99<br>(0.027)<br>** |
| Google query: "job" | 0 | **97.9%**<br>1.73<br>(0.044)<br>** | **95.9%**<br>1.57<br>(0.061)<br>* | **91.8%**<br>1.92<br>(0.032)<br>** |
| Google query: "employment" | 0 | **99.7%**<br>0.79<br>(0.217) | **98.0%**<br>2.18<br>(0.017)<br>** | **97.7%**<br>2.12<br>(0.021)<br>** |
| Google query: "job offers" | 0 | **96.0%**<br>1.76<br>(0.042)<br>** | **95.4%**<br>1.62<br>(0.056)<br>* | **93.9%**<br>1.70<br>(0.049)<br>** |
| Google query: "Labour office" | 2 | **99.1%**<br>1.23<br>(0.112) | **97.6%**<br>1.29<br>(0.101) | **97.9%**<br>1.10<br>(0.139) |
| Google query: "CV" | 4 | **96.8%**<br>1.75<br>(0.043)<br>** | **96.2%**<br>1.16<br>(0.126) | **95.0%**<br>1.24<br>(0.112) |

Source: author's calculations

Explanation: The percentage value shows the relative MSE of the augmented ARX model to the benchmark ARX model, value lower than 100% implies improved forecasts when additional variable was introduced. Next rows show the Clark-West test statistics, p-value and significance of this test. The null hypothesis is equal predictive accuracy of both models; we cannot reject it for high p-values, meaning that potential improvement is not statistically significant. Alternative hypothesis is superior forecasting quality of the larger model. Significance is denoted by asterisks (* for 10%, ** for 5%, and *** for 1% significance level).

offers" improve the ARX model even more than the AR(1) model alone, these improvements are statistically significant at 5% confidence level, and the total improvement increases in the subsamples in accordance with previous results. Query "employment" improves the benchmark only slightly (but significantly) and "CV" more but not statistically significantly. "Labour Office", however, does not improve the benchmark of ARX with Index of Industrial Production.

Similar conclusion for control variables is made based on Table (5.9), where the benchmark model is ARX with Composite Leading Indicators. Only

Index of Industrial Production and Consumer Confidence Indicator improve the benchmark, even though it is not always significant. On the other hand, Google data always improved the benchmark model. It is true for "job" and "job offers" as above, but also for "employment"; and "Labour office" also achieves better result.

Google search query data therefore improve not only the baseline AR(1) model, but also models already augmented with the best performing control variables, implying that it contains significant information compared with macroeconomic data. In this case, queries "job" and especially "job offers" are of a particular interest, since they had better results in combination with control variables than alone. For the whole sample, the best models of pairs of variables contained Index of Industrial Production together with "job offers", it improved the baseline AR(1) model by 14%. For subsamples, it was the query "job" together with Index of Industrial Production, improving baseline AR(1) by 11% for the second and 14% for the third subsample.

This conclusion – that Google data can significantly improve forecast also of benchmark models containing control variables – is in accordance with the related literature. For example D'Amuri & Marcucci (2010) also found for the U.S. unemployment that the best model always contained Google data, the same was stated by Chadwick & Sengul (2012) based on the analysis of Turkish data. The difference of our results compared with the related literature is the magnitude of these improvements.

For the U.S. data in the study of D'Amuri & Marcucci (2012), the best model containing Google data had the error lower by 18% compared with the best model without this kind of data. For Italian data, D'Amuri (2009) arrived at even bigger improvement, by 33%. And compared with baseline autoregressive process for Turkey (Chadwick & Sengul, 2012), models with Google data (only) had error smaller by 48% for nowcasts. In France, according to Fondeur & Karamé (2013), Google data improved the best model by 27% for a category of 15–24 years old, 18% for 25-49 and 10% for people older than 50 years. Choi & Varian (2012) found the improvement by 10% over the benchmark AR(1) process for the U.S. data, and Tuhkuri (2014) achieved the same result for Finland.

For the Czech data, these improvements are more humble, 10% being the maximum improvement when comparing the contribution of Google data in nested models. On the other hand, most of these improvements were statistically significant. It is also problematic to directly compare improvements across various studies. Firstly, researchers differ in the exact specifications of their models; some of the cited studies for example did not use first differences. Secondly, the quality of a benchmark is a determining factor, since it is easier to improve a worse model than a better one. In our case, even control variables were not able to improve the AR(1) model much, only by 7% in the best case.

### 5.3.3 Other results

So far, we have not commented on the results of Share of unemployed because it did not improve the benchmark models when measured by relative MSE (even though through the optics of the Clark-West test, which cleans for the effect of noise introduced when estimating larger model, the test statistics often suggested that model containing Share of unemployed was significantly better). This is probably because of the high correlation between the Share of unemployed and Unemployment Rate and its lags, meaning that the Share of unemployed does not bring new information into the autoregressive model already containing the first lag of the dependent variable.

In accordance with the nowcasting methodology applied, contemporaneous values of the Share of unemployed were not used since they are not known by the end of the month. Nevertheless, we also tested whether Google data bring additional information compared with the contemporaneous value of Share of unemployed (the results are presented in the Appendix B, Table B.1, in an analogous form to previously presented Tables). We found that except for the query "job offers", all other Google queries improved the benchmark model significantly at 5% confidence level – they bring additional information not captured by the data about registered unemployment.

## 5.4 Concluding remarks

We have shown that Google data can statistically significantly improve nowcasts of a baseline ARIMA(1,1,0) model for the unemployment rate. At the same time, this data significantly improve forecasts of models that contain best performing control variables, meaning they bring additional information not contained in other data. Some of the Google queries worked even better in combination with these control variables than alone.

At the same time, we have shown that Google data performed better when the first one or two years were neglected; Google data from years 2004 and 2005 contained noise that probably hampered the performance of the models estimated over this period. The improvement of Google data later in the out-of-sample part could also be caused by the increased popularity of both the use of the internet as a means of job search among unemployed people as well as increased popularity of Google search engine in the Czech Republic.

The best performing Google queries – ”job” and ”job offers” – were in the form of contemporaneous values. This is good for nowcasting, but cannot be used when forecasting for longer horizon. For this reason, we cannot say that Google data can predict unemployment in advance compared with control variables (usually, their $7^{th}$ or $8^{th}$ lag performed the best). But when nowcasting, for example for the out-of-sample period 2011–2013, query ”job” improved the baseline AR(1) process the most out of all analyzed variables, and for each subsample, the best model contained Google data. Therefore, Google data improves forecasts of macroeconomic and leading indicators. Similar findings for the Czech Republic were confirmed by Pavlíček & Krištoufek (2015), who also found that the incorporation of Google data improved their baseline models for unemployment nowcasting.

# 6 Consumer confidence

Confidence indicators have become closely followed in the past decade as one of the determinants or indicators of real economic activity. Dées & Brinca (2011) sum up that even when it is unclear whether the erosion of confidence was a cause or a consequence of the financial crisis after 2008, most of the academics agree that it prolonged and deepened this crisis and the following economic recession, having an impact on real economy.

Analysis of such interrelation between consumer confidence and real economic situation, such as the growth of GDP, is not new in the economic literature. For example, Giannone et al. (2009) and Bańbura et al. (2010) studied this relationship on the data from the Euro Area, de Bondt & Schiaffi (2011) from the Euro Area and the United States. Others also examined the relation between confidence and consumption, Ludvigson (2004) for the United States and Dées & Brinca (2011) for the Euro Area, among others.

For these purposes, authors usually use official indicators of confidence / sentiment.[19] Official indicators are usually based on questionnaires; these questions are sent (or asked telephonically) to managers / businessmen and to households. Managers are usually asked about their expected sales and orders in the near future; households are asked about their expectations about the situation in the whole economy as well as their own (mostly financial) situation. In general, it should provide information about the assessment and expectations of individual agents in the economy.

There is some critique of this process of information collection, summed up for example by Bechetti et al. (2012). Firstly, the answers are dependent on the exact formulation of the questions asked, but also on the current mood of the respondent; also, no real motivation exists to provide truthful answers, so the credibility may be a concern especially for sensitive questions.

Different proxies are also used for the sentiment of economic agents. Some studies use the assumption that people acquire most of the information from the media or that the level of attention the media pay to a particular phenomenon reflects the level of interest of people. For example Uhl (2011) analyzed the content of main American TV news broadcasts; Beber et al. (2013) also studied news releases in the U.S.

[19] These two terms are used as synonyms in this thesis.

Iselin & Silverstovs (2013) based their sentiment indicator on the occurrence of the word ”recession” in two German and Swiss newspapers and Tetlock (2007) performed a linguistic analysis on a popular column in the Wall Street Journal (concerned with stock markets) to create an indicator of investors pessimism. As Da et al. (2013) point out, the causality remains a question – whether the news drives the investors' sentiment, or whether it reflects it.

People use the internet for various purposes – they use it to communicate (e-mails, message boards, or social networks), they seek information of any kind (for example about goods and services), and they also gain knowledge about the current situation while reading online news. Some researchers started to use the internet as a source of information about sentiment, initially mostly in the area of investors; for example Antweiler & Frank (2004) examined the content of internet stock message boards; the potential of extraction of sentiment from Twitter is described in the literature review section.

By looking for any kind of information or news about particular phenomena on the internet using search engines, people reveal information about their concerns or even their current situation. For example by entering queries like ”recession” or ”credit card debt” (Da et al., 2013), ”debt burden” or ”energy costs” (Della Penna & Huang, 2009), people unknowingly reveal their sentiment (Wu & Brynjolfsson, 2013) Also, it is not likely that they would adjust their searching practice to manipulate an analysis of search query data, which may be a concern with questionnaires (Dzielinski et al., 2012).

Compared with methods based on questionnaires, search query data comes from a much larger number of people; costs of acquiring such data sample with regular methods would be immense (Hohhenstatt et al., 2011). But as Gill et al. (2012) mention, the representativeness of this sample to the whole population is a question; it differs from state to state based on internet penetration, internet skills, use of search engines and popularity of Google. The second hypothesis to be tested in our analysis is therefore:

> Search query data can be used to assess the sentiment of consumers in the Czech Republic in advance compared with other (e.g. survey-based) methods.

There are more ways how to deduce the sentiment and confidence of internet users based on volume of searches conducted with Google. Generally, one word, their combination, or a whole category of queries can be used.

Becchetti et al. (2012) analyzed a query "happiness" for Germany and Italy ("glück" and "felecita"), Da et al. (2013) used a whole group of words with economic and negative or positive meaning.

As summarized by Varian (2014), "*The challenge is that there are billions of queries so it is hard to determine exactly which queries are the most predictive for a particular purpose.*" In his previous article (Choi & Varian, 2012), he used a spike-and-slab regression to determine which of all categories of queries explain the Australian Roy Morgan Consumer Confidence Index – this Bayesian method provides a probability that a particular variable has a non-zero coefficient and belongs into the model.

Other authors chose their queries differently; Becchetti et al. (2012) chose the word "happiness" as an approximation for the life satisfaction of internet users (assuming that they rather search for what they do not have). In other cases, authors chose a larger group of categories in advance, and the final choice of queries was made based on the empirical relationship with the dependent variable.

Da et al. (2013) used the Harvard IV-4 dictionary to acquire the set of words that had and economic meaning and, at the same time, were in categories "positive" or "negative". They got a set of 149 words ("gold", "inflation", "depression", "crisis" among others), added top related searches from Google Trends, removed duplicates and entries with non-economic meaning (e.g. "depression medicine") and queries with incomplete time series.

With the final set of 118 queries, they made a regression against stock market returns, and the final sentiment index was an average of 30 queries that had the biggest correlation with returns (index created recursively every six month). They also confirmed the observation of Tetlock (2007) that in English language, negative terms had better explanatory power than positive ones, since all constituent parts of the final index had a negative correlation with market returns; they called it FEARS.

Beer et al. (2013) got inspired by this approach and conducted a similar analysis on French data. They extracted words with both economic and positive meaning (again using Harvard IV-4 dictionary), translated these 63 words to French and chose queries with a sufficiently complete time series. The final list consisted of 8 words (for example "crisis", "bankruptcy", "debtor" or "inflation") and the resulting index of negative sentiment was extracted using principal component analysis.

Della Penna & Huang (2009) examined the sentiment of internet users in the United States and other developed countries, but chose a different approach. They took categories they expected to be related to individual questions consumers are asked for the University of Michigan Consumer Sentiment Index (MCSI). Then, they regressed consumption variable against these categories and chose only those that had an expected sign when explaining current and future changes in consumption.

Their final index was an average over four categories: (a) bankruptcy – to approximate financial situation of a consumer; (b) luxury goods – to capture the willingness of households to spend on non-essential goods; (c) energy & utilities – as a concern of households with increasing prices; (d) office furniture – to capture the conditions in business. They motivated the choice of this category claiming that buying furniture is a non-essential long-term investment, purchased rather for new offices or to renew old ones when the conditions are good.

The above described articles looked at the sentiment from a general point of view – the financial situation of a customer as well as concerns about the economic situation. There is also a different approach to analyze consumers with the means of search query data that was also described in the literature review (Chapter 2) – looking directly at their consumption plans. Results of such articles will be also briefly commented on by the end of the discussion of application of sentiment indicators (Chapter 7).

When creating own index of consumers sentiment based on search query data, we got inspiration from both of the approaches described in detail above, taking into account characteristics of Google data in the Czech Republic. Firstly, categories of queries are not available for the Czech language, so only individual words or their combination can be used. Secondly, as has also been shown in the analysis of unemployment, a lot of the data are noisy and volatile during first few years of available data even for popular queries (with a complete time series); this is even bigger problem for less popular ones.

Initially, we wanted to follow the approach of Da et al. (2013). After extracting words with economic and negative meaning from Harvard IV-4 dictionary, translating them and downloading appropriate search volume data, most of the series were non-zero only during last few years, making this exact approach impossible to use. Some of the words, however, had sufficiently complete data to be used in further analysis. Nevertheless, we restricted the data sample to years 2007–2013 only (84 monthly observations), to allow all used time series to be complete and provide sufficient signal.

Finally, similarly to Da et al. (2013), we used individual words (or their combination in our case), and as in Della Penna & Huang (2009), the choice of relevant words was based on the questions asked in surveys for the official Consumer Confidence Indicator published by the Czech Statistical Office (CZSO). The description of this indicator (that will also serve as a baseline series for comparison of our sentiment index) is based on the information from the website of CZSO and also from the official guide of European Commission DG-ECFIN (2007).

During surveys for Consumer Confidence Indicator, respondents are asked questions about the following topics:

1) expected financial situation of household;
2) expected general economic situation;
3) expected unemployment (with inverted sign);
4) expected savings.

The questions are aimed towards the expected development over next 12 months. The answers are of a qualitative rather than quantitative nature: situation will get a lot better / a little better / no change / a little worse / a lot worse; number of unemployed will increase sharply / slightly / no change / fall slightly / sharply; saving money over next 12 months is very likely / fairly likely / not likely / not at all likely. The resulting balance is a difference between the number of positive and negative answers (stronger version having double the weights), expressed in percents.

## 6.1 Choice of components

An ideal procedure would be to use relevant categories, assign them to each of the questions and create the final index. Since this is not possible for Czech data, we used individual queries (of words or their combination) that may capture the essence of each of the questions from the point of view of internet users. These words serve as representatives of unobserved categories. Similar approach was used for example by Carrière-Swallow & Labbé (2010) for Chile when modeling automotive sales – because categories were not available, they used time series of most popular car brands.

We created a short preliminary list of queries potentially relevant for each question based. This list was based both on own ideas as well as suggestions taken from relevant literature. After that, a preliminary analysis was conducted: checking top related searches (using Google Trends) if the word is used in an appropriate context; examining its relation with the official Consumer Confidence Indicator; and a successful use of the query in the related literature was also taken into consideration. Paragraphs below describe queries included in the final analysis together with motivation for their choice for each of the questions.

- **Unemployment**

In accordance with our previous results, we chose queries that performed well when predicting the Czech unemployment rate, these are the following two:

  - □ ”job”
  - □ ”job offers”

- **Economic situation**

Based on the conclusions of Tetelock (2007) confirmed by Da et al. (2013) that negative words have better explanatory power in English, we tested this also for the Czech data. The first query related to this question was:

  - □ ”crisis” (”krize” in Czech)

We expect this word to capture the concerns of internet users with crisis of any kind, such as economic or financial, leading to lower sentiment level. The second word was chosen based on Della Penna & Huang (2009) who found that category ”office furniture” performed well for the US data. We tested this with a query:

  - □ ”furniture” (”nábytek” in Czech)

We extend their reasoning of a non-essential long-term investment also to households. New furniture can be bought when the old one is broken, but generally, this investment can be postponed or is done when equipping new household; we expect this query to be used more in a positive situation context.[20]

- **Financial situation**

Again, we used the assumption that words with negative connotation have better explanatory power, and also because it is less clear which positive words can be used in the context of the financial situation of a consumer (not describing consumption plans). Queries with financial meaning were used:

  - □ ”inflation + price increase” (in Czech, string ”inflace + zdražení + zdražování + zdražit”)

[20] This expectation is partially confirmed by correlation coefficients between GDP and an index of "Retail sale of furniture, electrical household appliances, hardware, paints and glass and other household articles in specialized stores". Over the period 2007–2013, the correlation was 0.57 between levels and 0.43 between quarterly growth rates of given variables.

This should cover people that are concerned with increased prices. To cover users already getting into financial trouble – already being close to bankruptcy or even bankrupt – we used the following:

- □ "distraint" (in Czech, string "exekuce + exekutor + exekutorský")

• **Savings**

The choice of a word is not obvious. For example the "interest" ("úrok" in Czech) may be searched both by people intending to save money, but also by those who intend to make borrowings. Also in general, it is not clear whether people saving money would indicate a bad situation – so called precautionary saving connected to low expectations, or a good situation – a person needs regular income higher than expenses to create savings. We used the following search query:

- □ "saving + savings" (in Czech, string "spoření + spořící")

In accordance with the official methodology, we used it with a positive sign in our analysis. We did not come up with the second query for the question about savings.

## 6.2 Index creation

To create the final index from the set of analyzed queries, we employed the procedure analogous to the creation of official indicators of confidence in Europe, when information from individual survey questions is aggregated into one index. We follow the description in DG-ECFIN (2007).[21] As mentioned above, the analysis of sentiment was conducted on the period 2007–2013 to allow Google data series to be complete, all the series were firstly shortened to this length. Then, the following procedure took place:

1) Google data series were seasonally adjusted using TRAMO/SEATS method.
2) All series were standardized:

$$\mathrm{Gn}_{it} = \frac{\mathrm{G}_{it} - \overline{\mathrm{G}}_i}{\sigma_{\mathrm{G}_i}} \qquad (6.1)$$

where $\overline{\mathrm{G}}_i$ is an average of Google data series $\mathrm{G}_i$ and $\sigma_{\mathrm{G}_i}$ is its standard deviation. The resulting series has a mean of 0 and a variance of 1.

[21] Page 20 of this document.

3) Series were aggregated into an index $Z_i$ using arithmetic average. All series were weighted equally; the only difference was in the sign given to each of the queries. Those with a positive meaning in our analysis remained with a plus sign (”furniture”, ”saving + savings”), queries with a negative meaning (the rest) were given a negative sign during the aggregation process.
4) The final index – Google Consumer Sentiment Index (GCSI) was calculated using the following formula:

$$\mathrm{GCSI} = \left(\frac{Z_i - \bar{Z}_i}{\sigma_{Z_i}}\right) * 10 + 100 \tag{6.2}$$

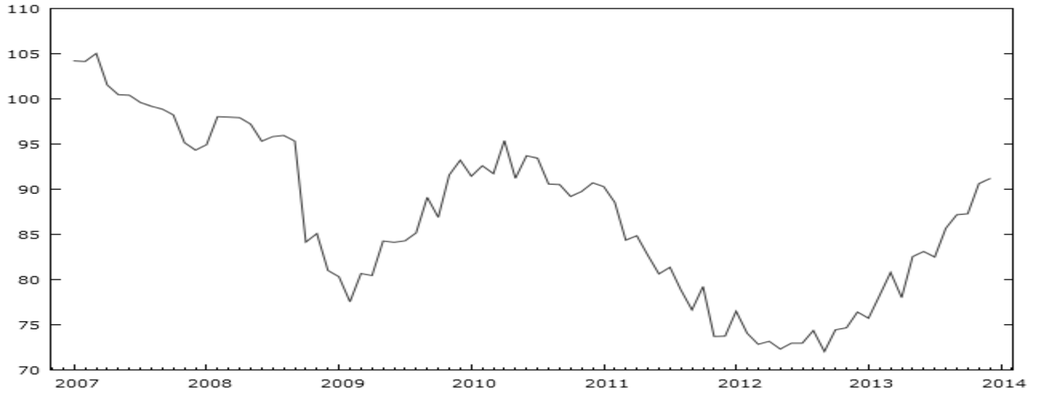
This means that the initial series $Z_i$ was standardized and adjusted so that its standard deviation is 10 and the mean is 100. Assuming normal distribution, the series would lie between 90 and 110 in approximately 68% of all cases.

The initial idea was to create the index based on different search queries and compare their performance. During our analysis, it appeared that the overall index – based on all 7 queries described above – performed reasonably well. Also, it has the best interpretation, since it includes two queries for every question (except for savings), and therefore aggregates more information. Especially for the questions about economic and financial situation, each of the queries captures different phenomenon (”crisis” x ”furniture”, ”inflation + price increase” x ”distraint”). The only disadvantage is that for the question about savings, only one query is used.

We called the final index Google Consumer Sentiment Index and denote it GCSI. It is thus an analogy to the official Consumer Confidence Indicator (CCI) of the Czech Statistical Office – it also measures positive sentiment (in contrast to some of the related studies). Figure (6.1) shows the development of the Czech Consumer Confidence Indicator over the period of 2007–2013.

The indicator was quite volatile over this short period – in the beginning of 2007, the consumer confidence started to fall. Around that time, first information about the problems in the U.S. real estate market started to emerge, but it is a question whether it directly influenced the confidence of Czech consumers. It kept falling until the end of 2008 and it remained at similar levels after that.

**Figure 6.1: Consumer Confidence Indicator (2007–2013)**




Source: CZSO

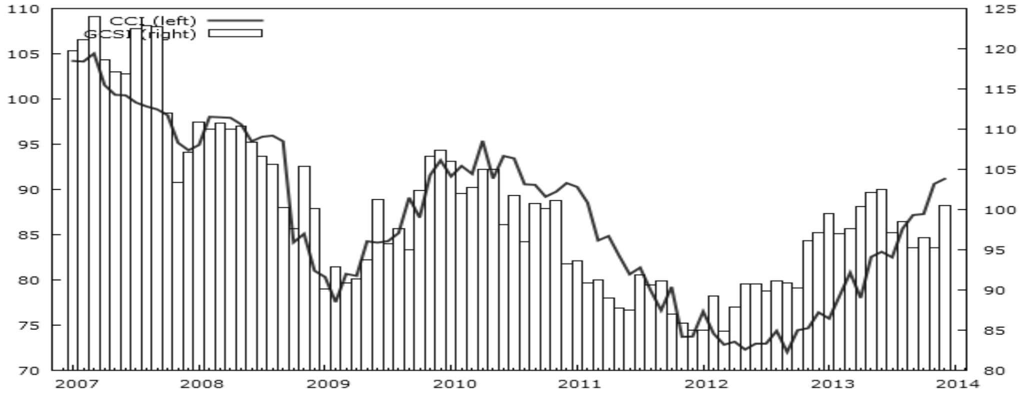
That changed in September 2008 with the fall of Lehman Brothers, and since then, the confidence dropped steeply to its local minimum in February 2009. From the point of Czech consumer, the situation started to improve after that, it was a time when European and other governments kept promising to stimulate the economy and proposed various economic packages (such as scrappage payments).

Next turn in the development of confidence came in the middle of 2010. At that time, the attention moved from the private to the public sector with the announcement of financial help (bailout loan) to Greece. The consumer confidence kept decreasing until the end of 2011, gradually achieving its minimum over the studied period. Since the middle of 2012, the confidence has been rising again.

## 6.3 Performance of GCSI

Figure (6.2) depicts the development of GCSI in comparison to the official indicator CCI. GCSI captures the development of CCI well – especially the turning points – during the first half of the analyzed period. Since then, GCSI

**Figure 6.2: Consumer Confidence Indicator and Google Consumer Sentiment Index**




Note: CCI – solid line; GCSI – bars. Source: CZSO, author's calculations starts to lead, CCI being lagged by several month; for example, GCSI started to fall sooner during the sovereign debt crisis, but also started to rise from the beginning of 2012 already.

To analyze this relationship formally, Table (6.1) provides correlation coefficients between levels of GCSI and CCI: between their contemporaneous values and their first lags; their significance is denoted by asterisks. The correlation coefficient between these two series is high, 0.85, which confirms visual observations; and the coefficient falls faster for the lagged value of CCI, indicating that GCSI is the leading of the two.

**Table 6.1: Correlation coefficients between Google Consumer Sentiment Index (GCSI) and Consumer Confidence Indicator (CCI)**

| | 1st lag of GCSI to current CCI | | Contemporaneous values | | 1st lag of CCI to current GCSI | |
|---|---|---|---|---|---|---|
| Correlation | 0.8445 | *** | 0.8535 | *** | 0.8034 | *** |

Source: author's calculations

Table (6.2) shows the correlations between logarithmic differences of both indicators. The only significant correlation coefficient is between the first lag of GCSI and the current value of CCI – 0.23, twice as much as the contemporaneous correlation.

**Table 6.2: Correlation coefficients between log-differences of Google Consumer Sentiment Index (GCSI) and Consumer Confidence Indicator (CCI)**

| | 1st lag of GCSI to current CCI | | Contemporaneous values | | 1st lag of CCI to current GCSI | |
|---|---|---|---|---|---|---|
| Correlation | 0.2321 | ** | 0.1091 | | 0.1451 | |

Source: author's calculations

Based on this analysis, it seems that GCSI leads CCI. To see if one index can explain changes in the other, we conducted OLS regressions – both for levels and log-differences; this was done again for contemporaneous values and first lags. The results are shown in Table (6.3). At first sight the results (significance of coefficients) copy the observations from correlation analysis. In this case, t-ratios are higher for lagged values of GCSI than for lagged CCI or even for contemporaneous values.

So far, the results are similar to the related literature – Da et al. (2013) found that the correlation between levels of the query "recession" and MCSI was -0.86 over the period 2004–2011 in the U.S.; the index of Della Penna & Huang (2009) had a correlation coefficient of 0.91 with MCSI, but the correlation was higher for contemporaneous values of log-differences (0.38) than for the first lag (0.29). The index of negative sentiment created by Beer et al. (2013) had also high correlation (-0.66) with the French confidence indicator.

**Table 6.3: Results of OLS regression (with constant) for levels and log-differences**

| **Levels** | 1st lag of GCSI to current CCI | | | Contemporaneous values | | | 1st lag of CCI to current GCSI | | |
|---|---|---|---|---|---|---|---|---|---|
| | T-ratio | P-value | Sign. | T-ratio | P-value | Sign. | T-ratio | P-value | Sign. |
| Results | 11.61 | (0.000) | *** | 11.59 | (0.000) | *** | 7.69 | (0.000) | *** |

| **Log-dif** | 1st lag of GCSI to current CCI | | | Contemporaneous values | | | 1st lag of CCI to current GCSI | | |
|---|---|---|---|---|---|---|---|---|---|
| | T-ratio | P-value | Sign. | T-ratio | P-value | Sign. | T-ratio | P-value | Sign. |
| Results | 2.30 | (0.024) | ** | 1.49 | (0.141) | | 0.95 | (0.344) | |

Source: author's calculations

Explanation: The table provides results of an OLS regression (with constant), only the estimates (t-ratio, p-value and significance) for analyzed variables are provided. Consumer Confidence Indicator (CCI) as a dependent and the first lag of Google Consumer Sentiment Index (GCSI) as an explanatory variable in the first column; contemporaneous values in the second column; and GCSI as a dependent and the first lag of CCI as explanatory variable in the third column. Robust standard errors (HAC) were used because of indications of autocorrelation and heteroskedasticity of residuals in some cases. Normality of residuals was rejected in some cases.

Values of t-ratios in OLS regressions are also similar to those in related studies for example Beer et al. (2013) arrived at t-ratio of 6.84 between lagged values of Google index and current values of official indicator for levels; and Da et al. (2013) had a t-ratio of 2.56 for log-differences of the Google query "recession", slightly higher than in our case. The analysis so far has indicated that GCSI can predict changes in CCI.

To formally test the hypothesis that our Google Consumer Sentiment Index leads the official Consumer Confidence Indicator, we conducted a Granger causality test. This test is described in the methodology section (Section 4.2.1), it is based on the estimates of Vector Autoregression (VAR) model for two (or more) variables. The null hypothesis of this test is a joint hypothesis that coefficients of all lagged values of one variable are equal to zero when explaining the second variable. This is tested using F-test and if we reject the null, we say that the first variable "Granger causes" the second variable; this relation may exists in one or both directions.

Even though generally it is not necessary for all series in the VAR system to by stationary – only the system to be stable as a whole – we need stationary time series for testing hypothesis within VAR. For levels of both time series, we could not reject the null of the presence of unit root using Augmented Dickey-Fuller (ADF) test and we also rejected stationarity using KPSS test. The opposite was true for log-differences – we rejected the presence of unit root with ADF test and did not reject stationarity with KPSS. For this reason, we used log-differences.

The correct specification of VAR model – appropriate number of lags included – is done using information criteria. AIC is usually recommended for monthly data, but it often overspecifies the model; BIC is recommended for quarterly data and smaller samples, HQIC for quarterly data. In our case, both BIC and HQIC suggested VAR(1) specification; AIC suggested VAR(3). At the same time, VAR(1) was the second model suggested by AIC, and VAR(3) by BIC and HQIC, while they had VAR(2) on the third place. For this reason, we conducted the Granger causality test for all three specifications for robustness reasons. The results are presented in Table (6.4).

**Table 6.4: Results of Granger causality test**

| | VAR(1) | | VAR(2) | | VAR(3) | |
|---|---|---|---|---|---|---|
| | GCSI =/=> CCI | CCI =/=> GCSI | GCSI =/=> CCI | CCI =/=> GCSI | GCSI =/=> CCI | CCI =/=> GCSI |
| P-value | 0.0139 ** | 0.2845 | 0.0681 * | 0.4106 | 0.0314 ** | 0.0116 ** |

Source: author's calculations

Explanation: The table provides p-values of Granger causality test based on VAR models with constant for log-differences of GCSI (Google Consumer Sentiment Index) and CCI (Consumer Confidence Indicator) for three different specifications. For each of this specification, p-values are provided for the test in both directions. The null hypothesis is that the first variables does not Granger cause the second one, we reject it for p-values lower than significance level; the significance is denoted by asterisks. Robust (HAC) standard errors were used because of heteroskedasticity indications.

In the VAR(1) model – which is similar to previous OLS estimations, only the effects of own lagged values is also taken into account – we reject the null hypothesis of no Granger causality only in the direction from GCSI to CCI, which is consistent with previous observations. The same is true for VAR(2) specification, suggesting that GCSI Grangers causes CCI. But for VAR(3), we reject the null in both directions, meaning that the Granger causality is present reciprocally, both indicators are closely interconnected.

## 6.4 Concluding remarks

In the analysis over the period 2007–2013, we have shown that Google data can be used to assess consumer sentiment in advance to survey-based methods. We achieved this by creating own index of consumer sentiment (GCSI) based on individual search queries relevant to questions asked in official surveys. GCSI captured turning points of the volatile development of the official indicator over this period; correlations coefficients suggested that GCSI was the leading of the two, which was confirmed also by Granger causality test (and one of the specifications suggested reciprocal relationship).

Queries used for our final index were chosen based on theoretical expectations, previous analysis (unemployment) and also performance in the related studies for other countries. By this, we have for example shown that the search query "furniture" can be a useful indicator even for Czech data, or that words with a negative connotation (or used in a negative context) have a good explanatory power also in the Czech language.

There are few caveats of our approach. Firstly, only one query was used for the question of savings. Secondly, the analysis was conducted only in-sample, creating the index at once having all the data from period 2007–2013, and testing the resulting index against the official one over this sample. The relationship found in-sample does not have to persist beyond December 2013.

When creating an index of consumer sentiment in real-time, one would proceed differently, with an approach analogous to recursive (or rolling) window scheme in forecasting. For example, the value of an index in January 2009 would be based on a data sample ending in January 2009, for February 2009 on a data sample ending in February 2009 and so on. Creating the index recursively and testing it out-of-sample is a suggestion for further analysis in this area of Google Econometrics.

In overall, we employed a simple and transparent approach, using few carefully chosen search queries (relevant to particular question) and aggregating them using official methods of statistical offices of the European Union. The resulting Google index performed well, but examining the correlation with the official confidence indicator is not the last step – we also need to test the contribution of GCSI when modeling real economic variables. If there was no connection, confidence indicators would not be of much use for economists. This is the subject of the next chapter.

# 7 Macroeconomic development

In the reviewed literature, authors chose various approaches to test the performance of their sentiment indicators. Da et al. (2013) tested their FEARS index on the stock market data for the United States, Beer et al. (2013) did the same for France. Becchetti et al. (2012) examined the relation between the query "happiness" and the spread of 10Y government bonds (a proxy for the threat of a financial crisis) for Germany and Italy, Della Penna & Huang (2009) modeled retails sales in the United States.

We decided to take a different approach in our analysis – to test our Google Consumer Sentiment Index to predict the overall economic situation. The use of Google data for this purpose is scarce; the only brief study we are aware of was conducted by Tkacz (2013), who modeled the probability of occurrence of a recession in Canada using Google query "recession". On the other hand, the use of official confidence indicators for short-term predictions (or nowcasts) of real economic activity is not new.

For example, de Bondt & Schiaffi (2011) analyzed the GDP growth in the U.S. and Euro Area and found that confidence indicators explained more variability in the data than the series of manufacturing plans. Giannone et al. (2009) and Bańbura et al. (2010) studied the Euro Area GDP and concluded that confidence indicators are better for short-term rather than long-term predictions. Household consumption expenditures in the U.S. and Euro Area were examined by Dées & Brinca (2011), they found that contribution of confidence indicators is bigger during more turbulent times; and Ludvigson (2004) modeled U.S. personal consumption expenditures finding modest but statistically significant improvements when using confidence indicators.

An interrelation between confidence and economic cycle is usually present, but as Ludvigson (2004) notes, the exact mechanism how households' attitudes influence the economy is a question. One of the most often quoted concerns about confidence indicators is to what extent it brings new information over fundamental economic variables, such as unemployment, inflation, interest rate, income, wealth, stock exchange and others. Ludvigson (2004) also raises a question whether such data contains information about future or rather about current and past events.

Beber et al. (2013) found indications for the U.S. data that confidence is based primarily on observations of the real economy. But generally, researchers find explanatory power of confidence over that of control variables; for example Ludvigson (2004) showed for the United States that confidence indicators bring more information than just expected changes in future income. Even if confidence data did not contain much new information over fundamental variables, it is not a problem if we want to use this data for forecasting; as Bańbura et al. (2010) notes, the usefulness of confidence data often comes from their early availability.

The last hypothesis to be tested in our thesis is the following:

> An analysis of search query data can help predict economic development and crisis in the Czech Republic more effectively compared with macroeconomic and financial data.

For the Czech Republic, Horváth (2012) studied the contribution of confidence indicators when forecasting GDP growth; Fišer (2010) examined the connection between Czech GDP and confidence indicators using Granger causality test. And Hermannová (2012) analyzed Czech confidence indicators in more detail, finding indications that they may bring useful information for nowcasting economic situation. Hermannová (2012) also took an inspiration from articles of Estrella & Mishkin (1998) and Taylor & McNabb (2007) – in addition to evaluating quantitative forecasts of economic growth, she tested the ability of confidence indicators to predict economic downturns as a binary variable.

We also took inspiration from this approach for the following reasons: all of the studies of GDP growth reviewed above were conducted using quarterly data. On the other hand, Google data are available with a higher frequency (so are confidence indicators) and we want to use this fact to our benefit. Having the information about economic situation in a binary form (0 – growth, 1 – downturn) makes it easier to convert these data to monthly frequency (either assigning appropriate values to all months of that quarter or using an external source of the data).

## 7.1 Economic downturns

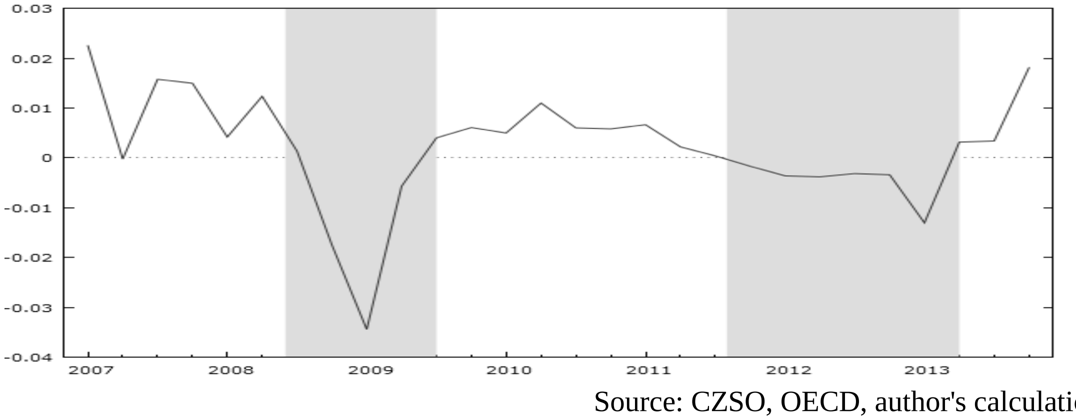
The Czech Statistical Office provides several definitions of an economic downturn / recession.[22] By the first definition, a technical recession occurs if there is a real decrease in quarterly GDP at least in two consecutive months. A wider definition says that both GDP and employment has to be reduced over

[22] http://www.czso.cz/eng/redakce.nsf/i/recession_depression

the same period. And a more general approach describes recession as a period between a peak and a trough of economic activity.

OECD (2014) provides the data about such peaks and troughs for its member countries, and because it is broken down to months (not only quarters), it is convenient for our analysis. We used this data for the Czech Republic, defining an economic downturn as a period between a peak (excluding) and a trough (including) over the period of 2007–2013. There were two such downturns between these years; Figure (7.1) depicts the development of quarterly GDP growth together with downturn periods (shaded area). GDP quarterly growth rate is calculated from a time series of seasonally adjusted Gross Domestic Product in constant prices of years 2005. Graphical analysis suggests that OECD peaks and trough data are derived from a below average (or negative) growth rate of real quarterly GDP for two or more periods.

**Figure 7.1: Quarterly growth rate of Czech GDP with downturn periods (2007–2013)**




Source: CZSO, OECD, author's calculations

### 7.1.1 Empirical strategy

To model dichotomous dependent variable, logit model (as described in the methodology section 4.4) is used. It is based on the assumption that while we observe a variable $y$ with only two possible values (0, 1), there is an underlying unobserved variable $y^*$ that determines when $y = 1$. Basically, we model the probability that $y = 1$. After the model is estimated, assuming $\hat{y} > 0.5$ is a prediction of $y = 1$, we can calculate the percentage of correct predictions. In addition, McFadden $R^2$ (based on values of likelihood functions of restricted and unrestricted model) is used to compare fits of models with different variables.

Models will be estimated over the period 2007–2013, using 84 observations of monthly data. Because of the length of the sample and characteristics of the dependent variable (2 downturn periods) we conducted only in-sample analysis. For an out-of-sample exercise, it is not possible to divide the sample in such a manner to allow the models to "learn" sufficiently from past downturns and to have sufficiently long out-of-sample period (containing both downturns and growth) for a good power of comparisons.

Our goal is to compare the explanatory power of Google data – our Google Consumer Sentiment Index – with the official indicators of confidence as well as with other explanatory variables often used in the related literature. Results are presented for the following variables:

- BCI ... Business Confidence Indicator
- CCI ... Consumer Confidence Indicator
- CLIs ... Composite Leading Indicators
- PX ... Prague Stock Exchange Index
- ER ... CZK / EUR Exchange Rate

### 7.1.2 Results

The results of individual models are presented in Tables (7.1) and (7.2). The first table provides results for lags 0 through 5 and the second table for lags 6 through 11, therefore showing the explanatory power for each variable towards economic downturn up to one year ahead. Each cell contains results for one model: individual variables are in columns and individual lags in rows of the presented matrix. Only the results important for interpretation and comparison of models are included: sign of beta estimate, significance of the model (compared with a restricted one), McFadden $R^2$ for model fit and the percentage of correct forecasts (measuring the model fit in a dichotomous framework).

Table (7.1) shows that for contemporaneous values and first few lags, Consumer Confidence Indicator performed the best among compared variables. The success of its predictions were around or above 80% even for up to 3 months ahead, meaning it could predict approximately 4 out of 5 downturn (and also growth) periods more than a quarter in advance. This superior performance is confirmed by McFadden $R^2$, CCI has better fit than other models. For contemporaneous value and first two lags also the Index of Prague Stock Exchange performed well, having even better predictions success that CCI, but it deteriorated faster with lagged values and its McFadden $R^2$ was lower. In overall, though, the PX index captures well the current economic situation or even the situation few months ahead.

**Table 7.1: Results of logit models for economic downturns (lags 0–5 for monthly data)**

| Lag | | GCSI | BCI | CCI | CLIs | PX | ER |
|---|---|---|---|---|---|---|---|
| 0 | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -44.9 *** | -47.9 *** | -32.5 *** | -37.5 *** | -39.4 *** | -55.0 ** |
| | McFadden $R^2$ | 0.213 | 0.160 | 0.430 | 0.343 | 0.310 | 0.037 |
| | % of correct forecasts | 70% | 68% | 85% | 77% | 85% | 58% |
| | % of correct downturn fc | 63% | 57% | 80% | 69% | 86% | 31% |
| 1 | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -45.4 *** | -50.8 *** | -36.1 *** | -42.6 *** | -43.4 *** | -53.7 ** |
| | McFadden $R^2$ | 0.197 | 0.100 | 0.361 | 0.246 | 0.232 | 0.049 |
| | % of correct forecasts | 70% | 61% | 84% | 71% | 83% | 61% |
| | % of correct downturn fc | 63% | 49% | 80% | 60% | 83% | 37% |
| 2 | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -44.4 *** | -53.0 ** | -38.6 *** | -47.0 *** | -46.9 *** | -51.8 *** |
| | McFadden $R^2$ | 0.206 | 0.052 | 0.311 | 0.161 | 0.161 | 0.074 |
| | % of correct forecasts | 72% | 59% | 82% | 68% | 78% | 68% |
| | % of correct downturn fc | 66% | 43% | 77% | 51% | 77% | 51% |
| 3 | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -44.7 *** | -54.2 | -41.2 *** | -50.3 *** | -49.2 *** | -49.5 *** |
| | McFadden $R^2$ | 0.192 | 0.021 | 0.256 | 0.092 | 0.112 | 0.107 |
| | % of correct forecasts | 73% | 52% | 82% | 59% | 75% | 73% |
| | % of correct downturn fc | 66% | 14% | 77% | 29% | 71% | 63% |
| 4 | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -45.0 *** | -54.6 | -44.5 *** | -52.5 ** | -50.7 *** | -46.7 *** |
| | McFadden $R^2$ | 0.180 | 0.004 | 0.188 | 0.042 | 0.075 | 0.149 |
| | % of correct forecasts | 74% | 59% | 76% | 55% | 71% | 75% |
| | % of correct downturn fc | 66% | 6% | 71% | 20% | 66% | 66% |
| 5 | Sign of Beta | - | + | - | - | - | - |
| | Log-likelihood | -46.0 *** | -54.2 | -46.5 *** | -53.7 | -51.8 ** | -44.1 *** |
| | McFadden $R^2$ | 0.152 | 0.001 | 0.143 | 0.010 | 0.046 | 0.187 |
| | % of correct forecasts | 72% | 56% | 70% | 53% | 66% | 73% |
| | % of correct downturn fc | 66% | 0% | 63% | 14% | 57% | 69% |

Source: author's calculations

Explanation: Results are presented for the set of 6 explanatory variables: GCSI (Google Consumer Sentiment Index), BCI (Business Confidence Indicator), CCI (Consumer Confidence Indicator), CLIs (Composite Leading Indicators), PX (Prague Stock Exchange Index), ER (CZK/EUR exchange rate).

Each cell contains results of one logit model – each column shows results for a particular explanatory variable, each row shows results for a particular lag of given variables when explaining current state of the economy (downturn).

Each cell contains the following: (1) sign of beta estimate: positive beta means that higher value of explanatory variable is connected with a higher probability of downturn (and vice versa); (2) log-likelihood and significance of the model based on likelihood ratio; (3) McFadden $R^2$ measuring fit of the model; (4) percentage of correct forecasts; (5) percentage of correct downturn forecasts (from all downturns).

Composite Leading Indicators by OECD, which is designed to capture turning points of economic situation, showed good performance for contemporaneous values, but it deteriorated significantly after the first lag. It is important to mention that compared with all other variables in this analysis, CLIs is published with a delay (approximately 1.5 months) and by the end of a particular month, only its second lag is known. This lag predicted only 51% of crises correctly during the analyzed period, so CLIs was not useful for nowcasting. Compared with other variables, Business Confidence Indicator also did not perform well even for contemporaneous values.

**Table 7.2: Results of logit models for economic downturns (lags 6–11 for monthly data)**

| Lag | | GCSI | BCI | CCI | CLIs | PX | ER |
|---|---|---|---|---|---|---|---|
| 6 | Sign of Beta | - | + | - | - | - | - |
| | Log-likelihood | -46.2 *** | -52.9 | -48.6 *** | -53.7 | -52.8 | -39.5 *** |
| | McFadden $R^2$ | **0.139** | **0.015** | **0.094** | **0.000** | **0.015** | **0.264** |
| | % of correct forecasts | 68% | 51% | 67% | 55% | 64% | 74% |
| | % of correct downturn fc | 63% | 23% | 57% | 0% | 46% | 71% |
| 7 | Sign of Beta | - | + | - | + | - | - |
| | Log-likelihood | -47.3 *** | -50.7 ** | -50.1 ** | -52.5 | -52.9 | -37.6 *** |
| | McFadden $R^2$ | **0.109** | **0.044** | **0.056** | **0.010** | **0.002** | **0.292** |
| | % of correct forecasts | 65% | 56% | 62% | 48% | 53% | 73% |
| | % of correct downturn fc | 60% | 34% | 51% | 20% | 0% | 71% |
| 8 | Sign of Beta | - | + | - | + | - | - |
| | Log-likelihood | -47.9 *** | -47.9 *** | -51.2 | -50.4 ** | -52.4 | -39.7 *** |
| | McFadden $R^2$ | **0.087** | **0.086** | **0.024** | **0.040** | **0.001** | **0.243** |
| | % of correct forecasts | 65% | 63% | 58% | 57% | 50% | 74% |
| | % of correct downturn fc | 60% | 49% | 46% | 46% | 9% | 74% |
| 9 | Sign of Beta | - | + | - | + | + | - |
| | Log-likelihood | -49.0 ** | -45.0 *** | -51.3 | -47.4 *** | -51.2 | -41.7 *** |
| | McFadden $R^2$ | **0.054** | **0.131** | **0.009** | **0.085** | **0.012** | **0.195** |
| | % of correct forecasts | 61% | 68% | 55% | 64% | 59% | 71% |
| | % of correct downturn fc | 57% | 57% | 40% | 60% | 34% | 71% |
| 10 | Sign of Beta | - | + | - | + | + | - |
| | Log-likelihood | -49.9 | -42.3 *** | -51.1 | -43.9 *** | -49.4 * | -43.3 *** |
| | McFadden $R^2$ | **0.024** | **0.173** | **0.001** | **0.143** | **0.035** | **0.154** |
| | % of correct forecasts | 61% | 72% | 53% | 68% | 61% | 70% |
| | % of correct downturn fc | 54% | 63% | 0% | 66% | 37% | 74% |
| 11 | Sign of Beta | - | + | + | + | + | - |
| | Log-likelihood | -50.1 | -40.0 *** | -50.5 | -40.0 *** | -47.6 ** | -44.9 *** |
| | McFadden $R^2$ | **0.008** | **0.208** | **0.000** | **0.209** | **0.059** | **0.111** |
| | % of correct forecasts | 60% | 71% | 52% | 74% | 62% | 70% |
| | % of correct downturn fc | 51% | 66% | 0% | 74% | 40% | 74% |

Source: author's calculations

Explanation: Results are presented for the set of 6 explanatory variables: GCSI (Google Consumer Sentiment Index), BCI (Business Confidence Indicator), CCI (Consumer Confidence Indicator), CLIs (Composite Leading Indicators), PX (Prague Stock Exchange Index), ER (CZK/EUR exchange rate).

Each cell contains results of one logit model – each column shows results for a particular explanatory variable, each row shows results for a particular lag of given variables when explaining current state of the economy (downturn).

Each cell contains the following: (1) sign of beta estimate: positive beta means that higher value of explanatory variable is connected with a higher probability of downturn (and vice versa); (2) log-likelihood and significance of the model based on likelihood ratio; (3) McFadden $R^2$ measuring fit of the model; (4) percentage of correct forecasts; (5) percentage of correct downturn forecasts (from all downturns).

The quality of Google Consumer Sentiment Index, on the other hand, was modest. Its $R^2$ was approximately half of the official confidence indicator for contemporaneous values and the first lag, and it predicted correctly two thirds of all cases. But its quality was consistent for more lags, compared with others whose fit and success in predictions deteriorated faster; for example for $4^{th}$ lag, the quality of GCSI was comparable to CCI. Also, the significance of its models is present even for $9^{th}$ lag. Table (7.2) shows the results for lags 6 through 11 to see the ability of variables to predict economic situation more than two quarters ahead.

Even though Exchange Rate did not perform well using contemporaneous values and first few lags, a significant feature of this variable is that its quality increases with additional lags. For example 5 months ahead, it was the best model measured both by $R^2$ fit as well as correct predictions, and this even rose further to the 7$^{th}$ lag. This is not the case for other variables, whose quality of models decay with the horizon of predictions, only our Google Consumer Sentiment Index had success of predictions higher than 50% for longer horizons.

Concerning the sign of beta coefficients estimates, for contemporaneous values, these are intuitively negative for confidence indicator, leading indicator and PX index. That means that a higher value of such indicator is connected with a smaller probability of downturn. This also seems to be true for the exchange rate, meaning that stronger CZK implies a higher probability of downturn in 7 months. For some variables, though, the sign reverts for longer horizon – most notably Business Confidence Indicator and Composite Leading Indicators – giving seemingly good fit with an unintuitive meaning – higher confidence increases the probability of a downturn in 11 months. This may be assigned to the regularity of downturns during the analyzed period.

To compare our results to the related literature, we estimated the models also on quarterly data. We converted all data series into quarterly frequency using averages over months of a particular quarter, obtaining time series of 28 observations over the sample 2007–2013. Downturn periods were also converted to quarterly data (based on the number of downturn months in that quarter). The results are presented in Appendix B (Table B.2).

The only difference observed using quarterly data compared with monthly data are absolute values of measures of fit, both McFadden $R^2$ and correct predictions. For example using contemporaneous values, CCI predicted correctly 90% of cases of both crisis and non-crisis periods, and Google data improved its success rate by 10 p.p. for current values and the first lag compared with individual months of this quarter; also its fit measured by $R^2$ improved significantly.

This observation confirms the claim of Estrella & Mishkin (1998) that qualitative results do not differ much when using monthly and quarterly data, only quantitative differences usually appear. According to them, pseudo-$R^2$ is smaller for monthly data because they are noisier; averaging them to quarters can remove some of the noise. This was confirmed by the Google index in our case, since it contains usually more noise than other series and indeed benefited from conversion to quarterly frequency the most.

Compared with Herrmannová (2012) who conducted similar analysis (only for confidence indicators and Composite Leading Indicators) for the Czech data for years 1999–2010, our results differ to some extent. She found that Business Confidence Indicator performed the best, which is in contrast with our findings; also in her case, the measure of fit of individual variables was lower compared with ours, for example correct number of predictions of downturns was only 50% for the best model for contemporaneous values – in our models, all BCI, CCI, and CLIs passed this threshold.

The results accord in two aspects. Similarly to us, Herrmannová (2012) also found the reversion of the sign of beta coefficient estimates for CLIs; implying that a better situation today (higher CLIs) translates to a higher probability of economic downturn in three quarters. This probably means a regular pattern in the Czech economic cycle longer than just over our analyzed period. Also, similarly to us, she concluded that confidence indicators are the most useful for nowcasting, since their contemporaneous values attain the best results.

Taylor & McNabb (2007) found good predictive power of confidence indicators even for four quarters ahead, studying the data over period 1983–1998 for the U.K, France (business confidence performed better in these countries), Italy and the Netherlands (consumer confidence performed better). From this regard, our results are similar to those of Italy and the Netherlands, but we did not find predictive power over such a long horizon. Concerning other compared variables, Estrella & Mishkin (1998) also showed, using U.S. data from 1959-1995, that stock exchanges indices have good predictive power even four quarters ahead in the United States. In our cases, it was up to one quarter.

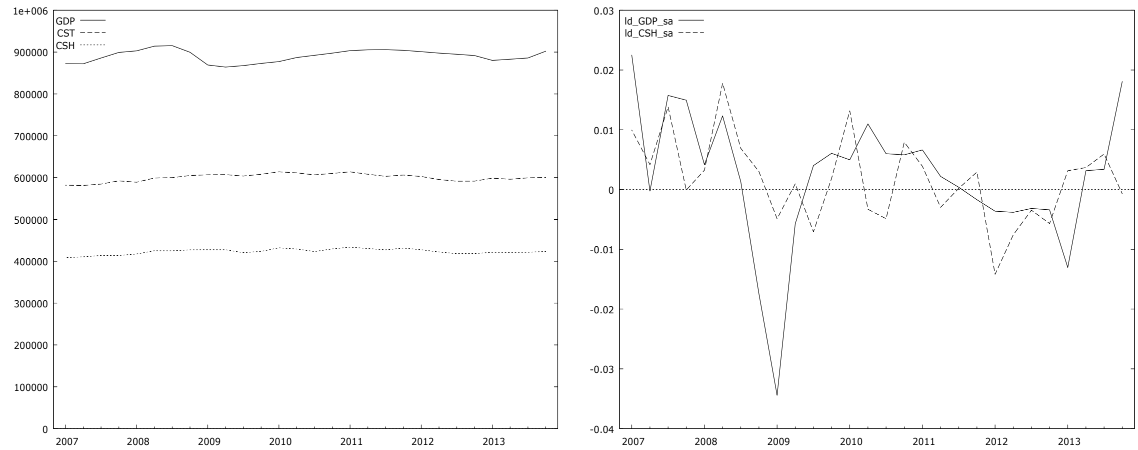
In one of the scarce studies of Google data in this framework, Tkacz (2013) used a search query "recession" in a probit model for the occurrence of recession in Canada on the data from 2004–2009. He concluded that this query would have been a good predictor of the recession of 2008/09 up to three months ahead, since it had the best in-sample fit for short horizons compared with the yield spread and credit, debit, and cheque payments data. In our case, Google data did not outperform control variables in short horizons, but belonged among the best for 4 and 5 months ahead.

## 7.2 Household consumption expenditures

Because we created a consumer sentiment index, we further extended our analysis to see how Google data performed when predicting consumption. To remain at the macroeconomic level, we used household final consumption expenditure – it is a component of Gross Domestic Product when calculated by expenditure approach, which is the sum of final use of goods and services by residential units (final consumption and the creation of gross capital) and the balance of exports and imports. The real final consumption is the sum of final consumption of households, the government, and non-profit institutions serving households; during the analyzed period (2007–2013), it accounted for around 66% of GDP.

Household consumption alone, whose data are based mainly on household budget surveys, accounts for almost half of the total GDP (around 47% from 2007–2013). Figure (7.2) shows the development of the GDP (solid line), total consumption (dashed line) and household consumption (dotted line) over the period 2007–2013. The share of consumption (both total and of households only) is relatively constant over time, but consumption did not fall significantly in 2008. This is confirmed by the RHS chart that depicts quarterly growth rates for GDP and household consumption.

**Figure 7.2: Czech GDP, total consumption and household consumption (2007–2013)**




GDP (solid line), total consumption (dashed line), and household consumption (dotted line), constant prices of year 2005.

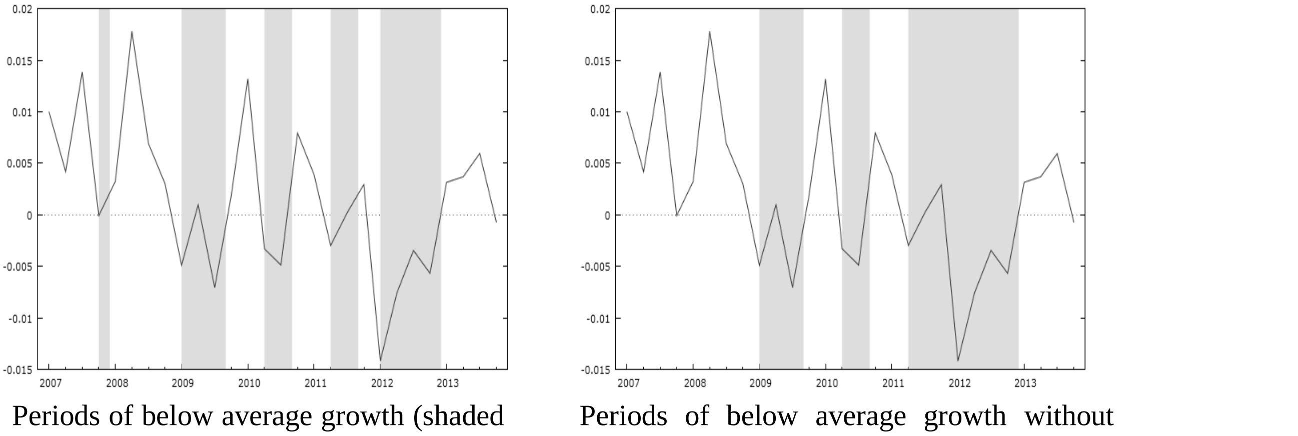
Quarterly growth rate of GDP (solid line) and household consumption (dashed line), constant prices of years 2005.

Source: CZSO, author's calculations

The growth rate of household consumption is more volatile than that of GDP, but the significant fall in GDP growth in the second half of 2008 was not caused by consumption. Given its nature as a constituent part of GDP, household consumption data are also published quarterly and with a significant delay. Even though the advantage of leading indicators is their early availability and higher frequency, it is hard to find a suitable variable to test their qualities.[23]

To test the predictive power of individual variables in the case of consumption, we followed the same approach as with economic downturns. Using quarterly data of household consumption, we calculated quarterly growth rate; the state of distress was defined as a below average growth of consumption, while any state had to last at least for two consecutive quarters to eliminate one-off changes.[24] This quarterly information was then converted to monthly frequency (assigned to respective months). Figure (7.3) shows periods of below average growth before and after this elimination.

**Figure 7.3: Quarterly growth rate of household consumption (2007–2013)**




Periods of below average growth (shaded area).

Periods of below average growth without one-off changes (shaded area).

Source: CZSO, author's calculations

In addition to Google Consumer Sentiment Index and explanatory variables used previously, we also wanted to capture the element of household income for predictions of its consumption. Unfortunately, similar indicators, such as average wage, are available only with quarterly frequency. As a proxy for household income in our analysis, we used the rate of unemployment (sea-

[23] Index of retail sales seems like a good candidate. However, for the analyzed period, we did not find a relation between this index and household consumption; based on predicting retails sales, it would not be possible to make conclusions about predictive abilities for household consumption.

[24] We conducted the analysis also without eliminating one-off changes, the results are included in the Appendix B (Tables B.4 and B.5); while quantitative measures differed to those presented below, qualitative conclusions made on such results were the same. This applies also to a possible definition of distress as a negative growth rate – the best model was superior to the control variables across all three specifications.

sonally adjusted) to assess the overall ability of households to spend money. The results are presented for the following variables:[25]

- BCI ... Business Confidence Indicator
- CCI ... Consumer Confidence Indicator
- CLIs ... Composite Leading Indicators
- PX ... Prague Stock Exchange Index
- U ... Unemployment Rate

### 7.2.1 Results

Tables (7.3) and (7.4) present the results of logit models for monthly data of below average consumption growth. Results indicate that our Google Consumer Sentiment Index performed the best among all other variables. This is true by the measure of fit using McFadden $R^2$ and also the percentage of correct predictions. GCSI predicted correctly 4 out of 5 cases using contemporaneous values or first lag, and still around 75% looking at longer horizons. These values are comparable to quality of CCI models for the whole economy.

Looking at control variables, Consumer Confidence Index was the best, but had significantly lower pseudo-$R^2$ than GCSI (even though percentage of correct prediction was comparable for contemporaneous values and first few lags). Therefore, we can say that our index of consumer sentiment is better suited for forecasting household consumption that the official indicator – CCI performed better modeling the overall economic situation than consumption itself (measuring by percentage of correct predictions, since $R^2$ cannot be compared between models for different variables); Google index had it the other way around.

BCI and CLIs are not good for consumption forecasting, these indicators predicted correctly less than 50% of distress periods even for contemporaneous values. On the other hand, PX index performed relatively well even for consumption, for example it had a better fit than CCI for the 3rd lag. The unemployment rate did not have a good fit measured by $R^2$, but the percentage of correct predictions was acceptable, for example almost two thirds of below average growth periods using contemporaneous values. Moreover, the sign of beta is intuitive as higher rate of unemployment is connected with higher probability of distress in consumption.

[25] Other variables, such as interest rate, monetary aggregates or households' credit were also tested. Because they did not perform better than presented variables, their results are not included.

**Table 7.3: Results of logit models for below-average growth of household consumption (lags 0–5 for monthly data)**

| Lag | | GCSI | | BCI | | CCI | | CLIs | | PX | | U | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Sign of Beta | - | | - | | - | | - | | - | | + | |
| | Log-likelihood | -33.0 | *** | -48.3 | *** | -38.9 | *** | -50.5 | *** | -45.0 | *** | -50.3 | *** |
| | McFadden $R^2$ | **0.425** | | **0.157** | | **0.322** | | **0.120** | | **0.216** | | **0.124** | |
| | % of correct forecasts | **82%** | | **56%** | | **77%** | | **62%** | | **68%** | | **57%** | |
| | % of correct downturn fc | **78%** | | **44%** | | **75%** | | **42%** | | **67%** | | **64%** | |
| 1 | Sign of Beta | - | | - | | - | | - | | - | | + | |
| | Log-likelihood | -33.1 | *** | -48.5 | *** | -38.8 | *** | -49.6 | *** | -44.0 | *** | -51.5 | *** |
| | McFadden $R^2$ | **0.418** | | **0.146** | | **0.316** | | **0.126** | | **0.225** | | **0.093** | |
| | % of correct forecasts | **82%** | | **55%** | | **80%** | | **58%** | | **69%** | | **55%** | |
| | % of correct downturn fc | **78%** | | **44%** | | **78%** | | **39%** | | **67%** | | **61%** | |
| 2 | Sign of Beta | - | | - | | - | | - | | - | | + | |
| | Log-likelihood | -34.3 | *** | -49.3 | *** | -41.3 | *** | -49.5 | *** | -42.9 | *** | -52.6 | *** |
| | McFadden $R^2$ | **0.390** | | **0.124** | | **0.265** | | **0.120** | | **0.237** | | **0.065** | |
| | % of correct forecasts | **79%** | | **55%** | | **77%** | | **55%** | | **70%** | | **52%** | |
| | % of correct downturn fc | **75%** | | **44%** | | **75%** | | **36%** | | **67%** | | **56%** | |
| 3 | Sign of Beta | - | | - | | - | | - | | - | | + | |
| | Log-likelihood | -35.7 | *** | -50.0 | *** | -42.9 | *** | -49.9 | *** | -42.4 | *** | -53.3 | ** |
| | McFadden $R^2$ | **0.358** | | **0.101** | | **0.228** | | **0.103** | | **0.238** | | **0.042** | |
| | % of correct forecasts | **75%** | | **52%** | | **74%** | | **54%** | | **68%** | | **49%** | |
| | % of correct downturn fc | **72%** | | **42%** | | **72%** | | **36%** | | **64%** | | **50%** | |
| 4 | Sign of Beta | - | | - | | - | | - | | - | | + | |
| | Log-likelihood | -36.9 | *** | -51.0 | *** | -46.2 | *** | -50.8 | *** | -44.1 | *** | -53.6 | * |
| | McFadden $R^2$ | **0.329** | | **0.074** | | **0.161** | | **0.078** | | **0.200** | | **0.027** | |
| | % of correct forecasts | **74%** | | **50%** | | **69%** | | **54%** | | **63%** | | **46%** | |
| | % of correct downturn fc | **72%** | | **39%** | | **67%** | | **36%** | | **58%** | | **42%** | |
| 5 | Sign of Beta | - | | - | | - | | - | | - | | + | |
| | Log-likelihood | -40.5 | *** | -51.5 | ** | -48.4 | *** | -51.6 | ** | -45.8 | *** | -53.6 | |
| | McFadden $R^2$ | **0.257** | | **0.054** | | **0.111** | | **0.052** | | **0.160** | | **0.015** | |
| | % of correct forecasts | **70%** | | **48%** | | **63%** | | **52%** | | **60%** | | **37%** | |
| | % of correct downturn fc | **69%** | | **36%** | | **61%** | | **28%** | | **56%** | | **17%** | |

Source: author's calculations

Explanation: Results are presented for the set of 6 explanatory variables: GCSI (Google Consumer Sentiment Index), BCI (Business Confidence Indicator), CCI (Consumer Confidence Indicator), CLIs (Composite Leading Indicators), PX (Prague Stock Exchange Index), U (Unemployment Rate).

Each cell contains results of one logit model – each column shows results for a particular explanatory variable, each row shows results for a particular lag of given variables when explaining current below-average growth of household consumption.

Each cell contains the following: (1) sign of beta estimate: positive beta means that higher value of explanatory variable is connected with a higher probability of downturn (and vice versa); (2) log-likelihood and significance of the model based on likelihood ratio; (3) McFadden $R^2$ measuring fit of the model; (4) percentage of correct forecasts; (5) percentage of correct downturn forecasts (from all downturns).

Looking at prediction for longer horizon in Table (7.4), only Google index GCSI preformed well, as it had statistically significant contribution even for three quarters ahead. This is in accordance with the results from economic downturns, where GCSI also had consistent performance across more lags compared with other variables.

In analogy with economic downturns, we conducted the same analysis also for household consumption on quarterly data; results are presented in Appendix B (Table B.3). Similarly to the overall economic situation, also in

**Table 7.4: Results of logit models for below-average growth of household consumption (lags 6–11 for monthly data)**

| Lag | | GCSI | BCI | CCI | CLIs | PX | U |
|---|---|---|---|---|---|---|---|
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -43.0 *** | -51.6 ** | -49.5 *** | -52.3 * | -47.1 *** | -53.4 |
| 6 | McFadden $R^2$ | 0.202 | 0.041 | 0.080 | 0.029 | 0.126 | 0.008 |
| | % of correct forecasts | 65% | 46% | 59% | 59% | 58% | 45% |
| | % of correct downturn fc | 64% | 33% | 56% | 28% | 56% | 17% |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -44.8 *** | -51.9 | -50.7 ** | -52.6 | -48.5 *** | -53.0 |
| 7 | McFadden $R^2$ | 0.158 | 0.024 | 0.047 | 0.012 | 0.089 | 0.004 |
| | % of correct forecasts | 61% | 42% | 55% | 57% | 55% | 56% |
| | % of correct downturn fc | 61% | 28% | 50% | 25% | 53% | 17% |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -46.1 *** | -52.0 | -51.4 | -52.4 | -50.3 ** | -52.5 |
| 8 | McFadden $R^2$ | 0.123 | 0.011 | 0.022 | 0.003 | 0.044 | 0.001 |
| | % of correct forecasts | 61% | 40% | 51% | 54% | 50% | 51% |
| | % of correct downturn fc | 61% | 25% | 39% | 19% | 47% | 3% |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -47.4 *** | -51.8 | -51.5 | -51.9 | -50.8 | -51.9 |
| 9 | McFadden $R^2$ | 0.087 | 0.003 | 0.007 | 0.000 | 0.023 | 0.000 |
| | % of correct forecasts | 57% | 41% | 47% | 52% | 48% | 52% |
| | % of correct downturn fc | 58% | 22% | 33% | 0% | 44% | 0% |
| | Sign of Beta | - | - | - | + | - | + |
| | Log-likelihood | -47.8 *** | -51.2 | -51.2 | -51.2 | -50.6 | -51.3 |
| 10 | McFadden $R^2$ | 0.067 | 0.000 | 0.001 | 0.001 | 0.014 | 0.000 |
| | % of correct forecasts | 58% | 51% | 41% | 38% | 47% | 51% |
| | % of correct downturn fc | 58% | 14% | 14% | 0% | 47% | 0% |
| | Sign of Beta | - | + | + | + | - | + |
| | Log-likelihood | -47.5 ** | -50.6 | -50.6 | -50.5 | -50.3 | -50.6 |
| 11 | McFadden $R^2$ | 0.060 | 0.000 | 0.000 | 0.002 | 0.006 | 0.000 |
| | % of correct forecasts | 58% | 37% | 44% | 59% | 47% | 48% |
| | % of correct downturn fc | 58% | 8% | 28% | 53% | 50% | 22% |

Source: author's calculations

Explanation: Results are presented for the set of 6 explanatory variables: GCSI (Google Consumer Sentiment Index), BCI (Business Confidence Indicator), CCI (Consumer Confidence Indicator), CLIs (Composite Leading Indicators), PX (Prague Stock Exchange Index), U (Unemployment Rate).

Each cell contains results of one logit model – each column shows results for a particular explanatory variable, each row shows results for a particular lag of given variables when explaining current below-average growth of household consumption.

Each cell contains the following: (1) sign of beta estimate: positive beta means that higher value of explanatory variable is connected with a higher probability of downturn (and vice versa); (2) log-likelihood and significance of the model based on likelihood ratio; (3) McFadden $R^2$ measuring fit of the model; (4) percentage of correct forecasts; (5) percentage of correct downturn forecasts (from all downturns).

this case, the conclusion made on quarterly data is identical. GCSI was the best across all lags among compared variables, the values of fit were even higher and percentage of correct predictions was more distinguishable from competing variables.[26]

In overall, when modeling the state of the economy as a binary variable, the official Consumer Confidence Indicator performed the best among

[26] For quarterly data, also information about average wage was available. We conducted the analysis using this data, but the results were not better compared to unemployment.

competing variables. Even though our GCSI is well correlated with CCI, it performed worse when modeling economic downturns. On the other hand, GCSI was superior to control variables (including CCI) when analyzing household consumption (which constitutes almost one half of the total GDP); we see it as a good indication of potential use of Google data for nowcasting or short-horizon forecasting of consumers' behavior.

## 7.3 Quantitative forecasts – VAR models

During the course of logit modeling, we converted the data into quarterly frequency. Because also data for GDP (and household final consumption expenditures) are available with this frequency, it allows us to test the predictive power of Google data also quantitatively, modeling growth rate of individual variables. In accordance with related literature, we decided to employ VAR models.

VAR models are used quite often when modeling GDP growth; for example by Giannone et al. (2009), Bańbura et al. (2010) or Taylor & McNabb (2007), to cite few from the related literature concerned also with confidence. Applying directly to Czech data, Horváth (2012) analyzed forecasting qualities of several variables using VAR model, similarly Herrmannová (2012). And when modeling household consumption, Dées & Brinca (2011) also used VAR among other models.

### 7.3.1 Empirical strategy

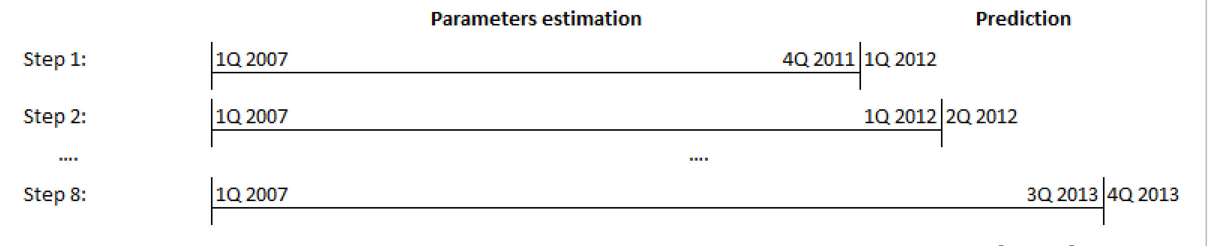
To test the potential contribution of individual variables in improved accuracy of predictions of GDP and consumption growth, we created out-of-sample forecasts for all models and compared their qualities. For nested models, the Clark-West test was used (in analogy to the analysis of unemployment); for non-nested models, we used the Modified-Diebold-Mariano test.

There were several issues to solve. Firstly, to test the contribution of individual variables – whether it contains any new information over macroeconomic or financial data – comparison of nested models is usually applied. For example for the Czech economy, Horváth (2012) describes that an appropriate baseline model for small open economies is a macro model containing inflation, interest rate (PRIBOR) and exchange rate (CZK/EUR).

In our case, using such large model is unattainable with a given dataset – 28 quarterly observations. For example in such macroeconomic VAR(2) model for GDP, containing one additional explanatory variable, there are 5 variables in total; therefore $(5 * 2 + 1) * 5 = 55$ coefficients to estimate in the whole system, with only 26 observations available in our dataset (loss of 2 observations for the lagged values).[27] For this reason, we firstly estimated only bivariate VAR models, and the quality of forecasts of individual models was compared using the Modified-Diebold-Mariano test for non-nested models. After that, models with one control variable were estimated and the quality of their forecast was compared by the Clark-West test.

Secondly, we need to divide the data into in-sample $R$ and out-of-sample $P$ parts. As is described in the methodological section, we generally want the out-of-sample period to be long enough to increase power of tests of predictive accuracy; on the other hand, we need a sufficient number of observations to estimate large number of coefficients. We divided the sample into $R = 20$ and $P = 8$, to have at least 20 observations for model estimation and 8 quarters to compare forecasts.[28] We used a recursive window scheme for out-of-sample forecasts creation to allow the models to use all available information given its already short sample size; this is described in Table (7.5).

**Table 7.5: Diagram of recursive window scheme for out-of-sample forecasting**




Source: author's description

Lastly, the lag length specification was set using Schwartz Bayesian Information Criterion. We chose VAR(2) model for GDP and VAR(1) model for household consumption.[29] Models were estimated using levels of additional explanatory variables, unless stated otherwise. Even if some of these were not

---

[27] Even in VAR(2) model for three variables, we need to estimate $(3 * 2 + 1) * 3 = 21$ coefficients, which is better, but still problematic for very short samples.

[28] We conducted the analysis also for division $R = 18, P = 10$ and $R = 16, P = 12$ to increase the ratio $R/P$ closer to 1 as recommended by Clark & McCracken (2011). Qualitative conclusions – especially for GCSI – did not differ much from those present in the analysis for $R = 20, P = 8$.

[29] In fact, BIC did not recommend this exact specification for all compared variables, but we chose only one specification for all variables to allow direct comparison of models. Still, even for variables for which BIC suggested different specification, VAR(2) for GDP and VAR(1) for consumption had the best forecasting performance.

stationary over the analyzed period, in the VAR framework, it is important that the whole model is stable, because cross-correlation is taken into account during estimation. Differencing the data is often discouraged (e.g. by Sims, 1980) to preserve long-term relationships between variables.

The procedure of forecast creation is analogous to ARX models in the unemployment analysis. In the reduced form of VAR(2) system, we have the following equation for the first variable (growth of GDP):

$$y_{1t} = a_{10} + a_{11}^1 y_{1,t-1} + a_{11}^2 y_{1,t-2} + a_{12}^1 y_{2,t-1} + a_{12}^2 y_{2,t-2} + e_{1t} \tag{7.1}$$

Therefore, the model is estimated using first $R$ observations, and these estimates of coefficient are used for forecasts:

$$\hat{y}_{1,t+1} = \hat{a}_{10} + \hat{a}_{11}^1 y_{1,t} + \hat{a}_{11}^2 y_{1,t-1} + \hat{a}_{12}^1 y_{2,t} + \hat{a}_{12}^2 y_{2,t-1} \tag{7.2}$$

This way, we obtained $P$ out of sample forecast, calculated forecast errors and Mean Squared Errors for each model. Equal forecast accuracy was tested either by the Modified Diebold-Mariano or Clark-West tests.

### 7.3.2 Results

Results of forecast comparison of non-nested models are presented in Tables (7.6) for quarterly GDP growth and (7.7) for quarterly growth of household consumption. Each table shows results of pairwise comparison of equal predictive accuracy between models of individual variables. The variables are ordered according to their MSE, the best model in the first row / column, the second best in the second row / column and so on. Then, relative MSE of the better model to the worse model is calculated and shown in the first row of each cell. The second row of each cell shows the Modified-Diebold-Mariano test statistic of equal predictive accuracy of both models with an alternative that the model with lower MSE is better; p-value for this statistics using Student $t_7$ distribution is in the third row and significant results are denoted with asterisks.

When modeling GDP growth, as is shown in Table (7.6), the results are very similar to the results from logit models. Consumer Confidence Indicator performed the best, MSE of its model was significantly better on the 10% confidence level than Composite Leading Indicators, Business Confidence Indicator and also Prague Stock Exchange Index. Actually, the PX index did not confirm its qualities from logit models when making quantitative forecasts (this may be caused by the fact that PX index performed the best for con-

**Table 7.6: Nowcasts of quarterly growth rate of GDP – pairwise comparison of non-nested models**

Comparison based on 8 out-of-sample forecasts (1Q 2012 – 4Q 2013)

| | CCI | ER | CLIs | **GCSI** | BCI | PX |
|---|---|---|---|---|---|---|
| CCI | | **83.3%**<br>0.33<br>(0.377) | **68.8%**<br>1.75<br>(0.062)<br>* | **68.5%**<br>0.76<br>(0.235) | **58.0%**<br>1.81<br>(0.057)<br>* | **51.8%**<br>1.50<br>(0.089)<br>* |
| ER | | | **82.5%**<br>0.35<br>(0.369) | **82.2%**<br>0.97<br>(0.182) | **69.6%**<br>0.89<br>(0.202) | **62.2%**<br>1.35<br>(0.110) |
| CLIs | | | | **99.6%**<br>0.01<br>(0.497) | **84.3%**<br>0.68<br>(0.259) | **75.4%**<br>0.64<br>(0.273) |
| **GCSI** | | | | | **84.7%**<br>0.50<br>(0.315) | **75.7%**<br>1.59<br>(0.078)<br>* |
| BCI | | | | | | **89.4%**<br>0.46<br>(0.331) |

Source: author's calculations

Explanation: The table shows the relative MSE when comparing each pair of models, name of each variable denotes VAR(2) model of GDP growth and this variable. Firstly, MSEs of forecasts of these models were calculated, and models were arranged from best to worst – the best model is in the first row / column, the second best in the second / column, etc.

Each cell shows the MSE of the better model relative to the worse one, making a pairwise comparison. The first row in each cell shows this relative MSE, the second row the Modified-Diebold-Mariano test statistic of equal predictive accuracy of both models with an alternative hypothesis that the model with lower MSE is better, p-value of this test using Student $t_7$ distribution; significance of the result is denoted by asterisks (* for 10%, ** for 5%, and *** for 1% confidence level).

Variables presented: Consumer Confidence Indicator (CCI); CZK/EUR Exchange Rate (ER); Composite Leading Indicators (CLIs); Google Consumer Sentiment Index (GCSI); Business Confidence Indicator (BCI); Prague Stock Exchange Index (PX).

temporaneous values in logit models, and VAR model uses only lags for forecasting). BCI index also did not perform well, which is in accordance with results from logit models. On the other hand, Exchange rate confirmed its qualities even for quantitative forecasts when using first two lags.

Google index also confirmed its qualities similarly to logit model – it was not the best, but performed better than BCI and PX and was almost equally good as CLIs. When compared with the official Consumer Confidence Indicator, even though it had the MSE less than 70% than GCSI, we cannot reject the null hypothesis of equal predictive accuracy. This is one of the disadvantages of testing forecast accuracy on a short out-of-sample, where even big differences in MSE are often statistically insignificant.

In logit modeling, the contribution of GCSI was greater when tested on household consumption. And this is true also for quantitative forecasts, as is shown in Table (7.7). In addition to previously used variables, we also present results for quarterly growth rate in average real wage (denoted Wage), because these data are available with quarterly frequency and there was no need to use other proxy (such as unemployment rate).

**Table 7.7: Nowcasts of quarterly growth rate of household consumption – pairwise comparison of non-nested models**

Comparison based on 8 out-of-sample forecasts (1Q 2012 - 4Q 2013)

| | **GCSI** | CCI | Wage | PX | CLIs | BCI |
|---|---|---|---|---|---|---|
| **GCSI** | | **74.2%**<br>0.73<br>(0.244) | **64.4%**<br>0.98<br>(0.180) | **60.3%**<br>2.58<br>(0.018)<br>** | **46.3%**<br>1.60<br>(0.077)<br>* | **45.3%**<br>1.98<br>(0.044)<br>** |
| CCI | | | **86.8%**<br>0.30<br>(0.387) | **81.3%**<br>0.53<br>(0.306) | **62.5%**<br>0.73<br>(0.245) | **61.1%**<br>0.86<br>(0.209) |
| Wage | | | | **93.7%**<br>0.18<br>(0.432) | **71.9%**<br>0.79<br>(0.228) | **70.4%**<br>0.89<br>(0.202) |
| PX | | | | | **76.8%**<br>0.82<br>(0.220) | **75.1%**<br>1.20<br>(0.135) |
| CLIs | | | | | | **97.9%**<br>0.22<br>(0.414) |

Source: author's calculations

Explanation: The table shows the relative MSE when comparing each pair of models, name of each variable denotes VAR(1) model of household consumption growth and this variable. Firstly, MSEs of forecasts of these models were calculated, and models were arranged from best to worst – the best model is in the first row / column, the second best in the second / column, etc.

Each cell shows the MSE of the better model relative to the worse one, making a pairwise comparison. The first row in each cell shows this relative MSE, the second row the Modified-Diebold-Mariano test statistic of equal predictive accuracy of both models with an alternative hypothesis that the model with lower MSE is better, p-value of this test using Student $t_7$ distribution; significance of the result is denoted by asterisks (* for 10%, ** for 5%, and *** for 1% confidence level).

Variables presented: Google Consumer Sentiment Index (GCSI); Consumer Confidence Indicator (CCI); quarterly growth rate of average real wage (Wage); Prague Stock Exchange Index (PX); Composite Leading Indicators (CLIs); Business Confidence Indicator (BCI).

For household consumption, GCSI was the best among compared variables; this difference was statistically significant compared with PX index, CLIs and BCI. The improvement of forecasts over Wage was also large (65%) as well as over CCI (75%), but this was statistically insignificant. In overall, this confirms observations made on logit models where GCSI was the best and CCI the second one.

Also results for CLIs and BCI are not surprising given their performance in logit models. On the other hand, PX index performed better for GDP than for consumption in logit models, and it is the opposite here. This is also caused by the fact than even for logit, lagged value performed better for consumption than for GDP, so the results are consistent. Generally, though, we can see that even when there are large differences in MSE, we reject the null of equal predictive accuracy only in some cases for GCSI.

To check whether Google data brings additional information compared with other variables, we also tested forecasts accuracy for nested models using the Clark-West test; this procedure was analogous to that used in unemployment analysis. This time, the benchmark (smaller model) is a bivariate VAR for GDP growth / consumption growth and a control variable; the bigger model is a trivariate VAR including one additional variable in the system.

For models of GDP growth, we used Composite Leading Indicators in the benchmark model. This choice is motivated by the fact that it is based on a set of leading indicators that were found useful for the Czech economy by OECD, including macroeconomic variables (but for at least a part of the analyzed period, it contained also information about CCI and PX index).

For models of consumption growth, we used the growth of average real wage in the benchmark model. This is because average wage can serve as a proxy for household income, which in theory should be connected to household consumption. It also performed well in the comparison of non-nested models, being for example superior to Composite Leading Indicators.

The results are presented in Table (7.8), both for GDP growth and consumption growth. The table shows the relative MSE of models augmented with additional variables compared with the benchmark model. The left column shows results for GDP growth, the right column for consumption growth. Relative MSE (%) is accompanied with the Clark-West test statistic of equal predictive accuracy and its p-value (using Student $t_7$ distribution) with an alternative hypothesis that the bigger model has better forecasts.

For models of GDP growth, all control variables improved the benchmark model. This improvement was statistically significant at 5% confidence level for Consumer Confidence Indicator, Exchange Rate (who performed better also in comparison of non-nested models), but also Google Consumer Sentiment Index. Therefore, Google data contained significant additional information compared with macroeconomic variables represented by CLIs.

In addition, even though GCSI performed equally well as CLIs in a bivariate framework, the resulting MSE was less than 60% when these variables were combined together. This model was also the best among compared models, showing that both variables work well together. This is not true for example for CCI – its MSE relative to CLIs was 68.8% in a bivariate setting and 65.5% when combined in one model together – only small improvement, which may be caused by the fact that since the 2[nd] quarter of 2012, CLIs contains also information about CCI.

**Table 7.8: Nowcasts of quarterly growth of GDP and household consumption – comparison of nested models**

Comparison based on 8 out-of-sample forecasts (1Q 2012 - 4Q 2013)

| | GDP | | Consumption |
|---|---|---|---|
| Benchmark model: | VAR(2) with CLIs | | VAR(1) with Wage |
| MSE of the benchmark model: | 0.0000685 | | 0.0000483 |
| Additional variable in the system | % MSE of benchmark | | % MSE of benchmark |
| Google Consumer Sentiment Index (GCSI) | **58.2%**<br>2.85<br>(0.012)<br>** | | **66.0%**<br>1.63<br>(0.074)<br>* |
| Business Confidence Indicator (BCI) | **94.5%**<br>1.07<br>(0.159) | | **120.5%**<br>-1.16<br>---- |
| Consumer Confidence Indicator (CCI) | **65.5%**<br>2.67<br>(0.016)<br>** | | **92.7%**<br>0.97<br>(0.183) |
| Composite Leading Indicators (CLIs) | **-**<br>-<br>- | | **117.2%**<br>-1.85<br>---- |
| Prague Stock Exchange Index (PX) | **91.0%**<br>1.41<br>(0.101) | | **99.3%**<br>0.49<br>(0.318) |
| Echange Rate CZK/EUR (ER) | **63.9%**<br>2.97<br>(0.010)<br>** | | **-**<br>-<br>- |

Source: author's calculations

Explanation: The table displays results of comparisons of nested models. The benchmark model for GDP growth (left column) is VAR(2) model for GDP growth and Composite Leading Indicators (CLIs), the benchmark for consumption growth (right column) is VAR(1) model for consumption growth and growth of average real wage (Wage). Results in individual cells show MSE of a model augmented with variable described in a given row relative to the MSE of the benchmark model. In each cell, the results show relative MSE; the Clark-West test statistic of equal forecasting accuracy with an alternative hypothesis of superior forecasting quality of the larger model; p-value of the CW statistic using Student $t_7$ distribution; significance is denoted by asterisks (* for 10%, ** for 5% and *** for 1% significance level).

The results of consumption growth show that GCSI is the only variable improving the benchmark (bivariate VAR with consumption growth and growth of average real wage), again showing that it contains significant additional information. On the other hand, the improvement when using CCI was only small and almost non-existent when using PX. CLIs and BCI did not improve the benchmark at all.

But when looking at actual relative MSE, and comparing with non-nested models, we can notice one contrasting thing: for the consumption growth, we would be better off using bivariate models instead of trivariate ones. For example, a bivariate model of GCSI had MSE relative to Wage 64.4%; when combined together into trivariate VAR, this relative MSE is 66%. And it is similar for CCI, relative MSE is 86.8% for non-nested models and 92.7% for nested ones. Interestingly, the same conclusion would be made if Composite Leading Indicators was used as a benchmark for consumption growth. This partially confirms the claim of Varian (2014) that "*simpler methods tend to work better for out-of-sample analysis*".

## 7.4 Concluding remarks

In overall, the analysis of VAR modeling and forecasting confirmed most of the conclusions made based on logit models. Google Consumer Sentiment Index performed better (relative to other variables) when modeling household consumption than the GDP itself. For household consumption, it performed the best among the whole set of control variables.[30] For GDP, it performed better than Business Confidence Indicators and equally well as Composite Leading Indicators. And when combined with CLIs, such trivariate VAR provided forecasts better than others.

Also conclusions about Consumer Confidence Indicator from logit model were confirmed for out-of-sample nowcasting in VAR setting. It performed the best among competing variables for GDP and was second after GCSI when modeling household consumption. Exchange rate confirmed its qualities even for point forecasts, and CLIs performed better for GDP than for consumption. BCI, on the other hand, did not work well in either case.

This is consistent with the results of Herrmannová (2012), she also found that for out-of-sample forecasts of Czech GDP growth, CCI performed better compared with CLIs and BCI (when comparing non-nested models,

[30] For some of them, the analysis was conducted but results were not presented. For example inflation, unemployment, monetary aggregates, interest rate.

making one-, two-, three- and four-steps ahead forecasts at once); also, CCI was the only one (of those three indicators) that significantly improved the macro model for Czech economy (inclusion of inflation, interest rate and exchange rate). On the other hand, Horváth (2012) did not find significant contribution of CCI to macro model's forecast accuracy; and Fišer (2010) found Granger causality between confidence and economic development only in particular model specifications (and a stronger relation in the other direction).

Concerning specifically Google data, an analysis of their performance when explaining overall economic situation is not that common. Individual studies rather tested the interrelation between sentiment indicators (or consumption plans) based on Google data and private or retail consumption. For this reason, direct comparison of quantitative improvements of forecasts is not possible.

In overall, the authors of such studies found that Google data contain useful information about spending plans of consumers. This was discovered using different models specification and different ways of extracting information from Google data (using individual queries, whole categories, principal components, etc.). Often, such data provide better fit or forecasts compared with official confidence indicators, as concluded for example by Kholodilin et al. (2010) and Della Penna & Huang (2009) for the U.S. or Schmidt & Vosen (2012) for Germany. We found the same even for the Czech Republic.

Others, like Suhoy (2009) for Isreael, Carrière-Swallow & Labbé (2010) for Chile, or Schmidt & Vosen (2011) for the U.S., also arrived at the conclusion that compared to control variables, Google data captured well turning points in consumption development. This is something we observed also for our GCSI index which captured well beginnings of periods of below average growth of household consumption.[31]

[31] A graphical representation is provided in Appendix B, Figure (B.2).

# 8 One year later: Results revisited

Since the internet is one of rapidly evolving areas, the question is whether internet based data can be used stably over time. For the purpose of this rigorous thesis, we have decided to test it and conduct the analysis on a dataset extended by one year. That is, each time series used in the original study was prolonged by 12 observations for monthly data and 4 observations for quarterly data.[32] Only GDP and household consumption time series were substituted by new data – due to a change in methodology by CZSO in accordance with ESA 2010 standard, the data for year 2014 were inconsistent with the original series.[33] This chapter summarizes main finding of the analysis identical to the above for each area of research.[34]

## 8.1 Unemployment

We used the AR(2) model for the log-differences of the unemployment rate for the analysis of the series on a data sample from 2004 to 2014. This model was chosen based on the Box-Jenkins methodology and at the same time, this model alone and enriched with additional variables performed better compared with AR(1) in terms of forecasting accuracy. Out-of-sample forecasts were made using a rolling window of 60 observations for parameters estimations. The results are presented in Appendix C (Tables C.1, C.2, C.3) in an identical form to Chapter 5.

Similarly to the original case of out-of-sample period of 2009-2013, the improvement of forecasts, when augmenting the baseline AR(2) process by additional explanatory variables, was only modest when looking at the 2009-2014 out-of-sample period. The only statistically significant improvements were 6.5% for the Index of Industrial Production and 3.3% for the Google query "job offers".

The modesty of improvements was apparent even for subsamples of 2010-2014 and 2011-2014, indicating either poor quality of explanatory variables or a good quality of a pure AR(2) process for log-differences of the un-

[32] In case the original time series was revised by the publishing authority, these changes were not taken into account, only new observations for the year 2014 were added.

[33] http://apl.czso.cz/pll/rocenka/rocenka.avizo_revize?mylang=EN

[34] All necessary tests including stationarity, stability, and others were conducted exactly the same as in the original thesis, but results are reported only for the main findings for each topic.

employment rate. The only statistically significant improvements, by 4% and 5% respectively, were recorded by "job offers". Even when not statistically significant, search queries "job" and "CV" also slightly improved the AR(2) forecasts on both subsamples, similarly to Index of Industrial Production and the Composite Leading Indicators.

The strength of Google data proved again in a combination with other explanatory variables. Even though most often not statistically significant (sometimes only slightly), search queries "job" and "CV" recorded improvements of both baseline models – one with the Index of Industrial Production and one with the Composite Leading Indicators – greater than of the AR(2) model alone, and also performed superior to other control variables.

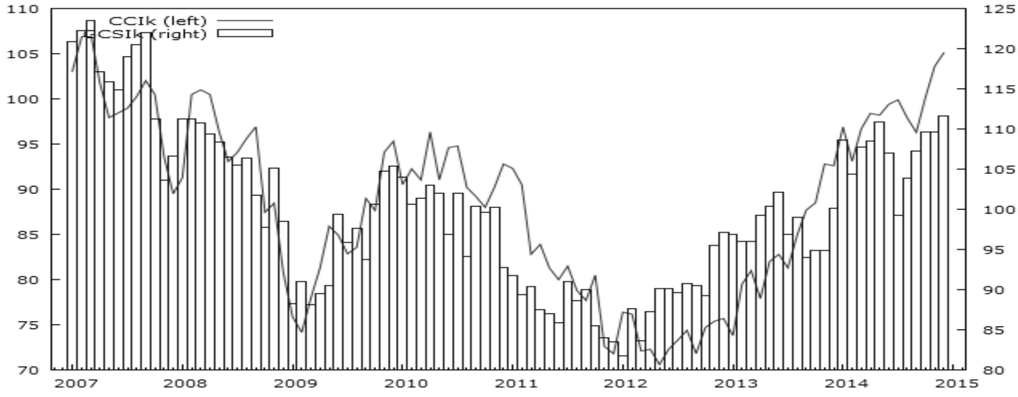
Most importantly, Google query "job offers" recorded statistically significant improvements of both baseline models in all three subsamples, and was also contained in the best model for each of the subsamples when comparing with the baseline AR(2) process. These best models were its combinations with Index of Industrial production for out-of-sample periods 2009-2014 and 2010-2014 (improvements of AR(2) forecasts by 10% and 6% respectively) and with Composite Leading Indicators for the period 2011-2014 (improvement by 8%).

In overall, while the ability of individual explanatory variables to improve the baseline autoregressive process forecasts decreased, Google data, especially the query "job offers", performed well and proved useful especially in combination with best performing control variables. This was also true for the queries "job" and "CV", which indicates that Google data contains additional information compared with macroeconomic variables. This is especially relevant regarding the queries "job" and "job offers", since we originally included them in the creation of our Google Consumer Sentiment Index, and we can continue to do so.

## 8.2 Consumer confidence

We created the Google Consumer Sentiment Index (GCSI) the same way as described in Chapter 6, using the same Google queries as well as the methodology applied, only this time on a data sample 2007-2014. Figure (8.1) depicts the development of both official Consumer Confidence Indicator of the Czech Statistical Office and GCSI over this period.

Figure 8.1: **Consumer Confidence Indicator and Google Consumer Sentiment Index**




Note: CCI – solid line; GCSI – bars. Source: CZSO, author's calculations

During the year 2014, by a visual analysis, both series developed quite similarly. Both grew on average over the year and both recorded two drops – one smaller in February, lasting one month, and one longer in the mid-year, lasting two consecutive months. One of possible explanations for both drops may be the crisis in the eastern Ukraine, with a start of so the called Crimean crisis in February.

By the mid-2014, as a response to a wave of various economic and non-economic sanctions imposed on Russia by the United States, the European Union and other countries, Russia announced an embargo on imports of most the agricultural and some other products from various countries, including the Czech Republic. These measures and general deterioration of commercial relationships with Russia were vastly discussed to have a negative impact on slowly recovering European economies, and could also influence the sentiment of Czech consumers as reflected by the indices.

After that, both indices grew and by the end of 2014, both achieved their highest values since 2008. This interconnection is in accordance with the original analysis over the sample 2007-2013, and GCSI again slightly appears to be the leading one. To analyze the interconnection of both series formally, Table (8.1) summarizes correlation coefficients between levels of GCSI and CCI – their contemporaneous values and their first lags, with significances denoted by asterisks.

**Table 8.1: Correlation coefficients between Google Consumer Sentiment Index (GCSI) and Consumer Confidence Indicator (CCI)**

| | 1st lag of GCSI to current CCI | | Contemporaneous values | | 1st lag of CCI to current GCSI | |
|---|---|---|---|---|---|---|
| Correlation | 0.8247 | *** | 0.8473 | *** | 0.7816 | *** |

Source: author's calculations

The correlation coefficients are almost identical to the original analysis, being the largest for contemporaneous values, and GCSI being potentially the leading of the two. This is also confirmed by correlations analysis performed on log-differences presented in Table (8.2). Compared with the original analysis, the correlation coefficient for contemporaneous values increased significantly (more than doubled), while the ones for the first lags decreased.

**Table 8.2: Correlation coefficients between log-differences of Google Consumer Sentiment Index (GCSI) and Consumer Confidence Indicator (CCI)**

| | 1st lag of GCSI to current CCI | Contemporaneous values | 1st lag of CCI to current GCSI |
|---|---|---|---|
| Correlation | 0.1815 * | 0.2329 ** | 0.0873 |

Source: author's calculations

The tighter connection between contemporaneous values of log-differences rather than for lags of both indices is also confirmed by the Granger causality test, the results are presented in Table (8.3). In all three cases, we cannot reject the null hypothesis that coefficients of all lagged values of one variable are equal to zero when explaining the second variable, so we cannot say that the first "Granger causes" the other. On the other hand, this non-rejection was very close for the VAR(1) specification (which was suggested by all three information criteria as the best specification) in the direction from GCSI to CCI.[35]

**Table 8.3: Results of Granger causality test**

| | VAR(1) | | VAR(2) | | VAR(3) | |
|---|---|---|---|---|---|---|
| | GCSI =/=> CCI | CCI =/=> GCSI | GCSI =/=> CCI | CCI =/=> GCSI | GCSI =/=> CCI | CCI =/=> GCSI |
| P-value | 0.1022 | 0.2137 | 0.2196 | 0.4140 | 0.2114 | 0.2116 |

Source: author's calculations

Explanation: The table provides p-values of Granger causality test based on VAR models with constant for log-differences of GCSI (Google Consumer Sentiment Index) and CCI (Consumer Confidence Indicator) for three different specifications. For each of this specification, p-values are provided for the test in both directions. The null hypothesis is that the first variables does not Granger cause the second one, we reject it for p-values lower than significance level; the significance is denoted by asterisks. Robust (HAC) standard errors were used because of heteroskedasticity indications.

Therefore, the analysis has shown that the interconnection between both indices has persisted even beyond December 2013, using originally chosen, quite diversified Google queries and aggregating them in accordance with the original procedure. The development of both indices appeared to be close in year 2014, the correlation coefficients for contemporaneous values was the highest, while GCSI appeared to be the leading of the two indices. We could not reject the null hypothesis of no-Granger causality only by a small margin on the 10% significance level. This relative stability of results is encouraging and allows us to test the quality of GCSI when predicting overall economic development.

---

[35] It should be noted that when analyzing the period 2007-2014, TRAMO/SEAT analysis in Gretl, in contrast to the 2007-2013 sample, didn't identify and remove any seasonality from the CCI series. This could have influenced both the correlation coefficients and the Granger causality test results.

## 8.3 Macroeconomic development

For the macroeconomic development, we tested the change in quality of GCSI both in logit models (economic downturns, below average growth of household consumption) and quantitative forecasts using VAR models for GDP and household consumption, over the period 2007-2014. Extensive results are included in the Appendix C, and only a summary of the most important results is provided below.

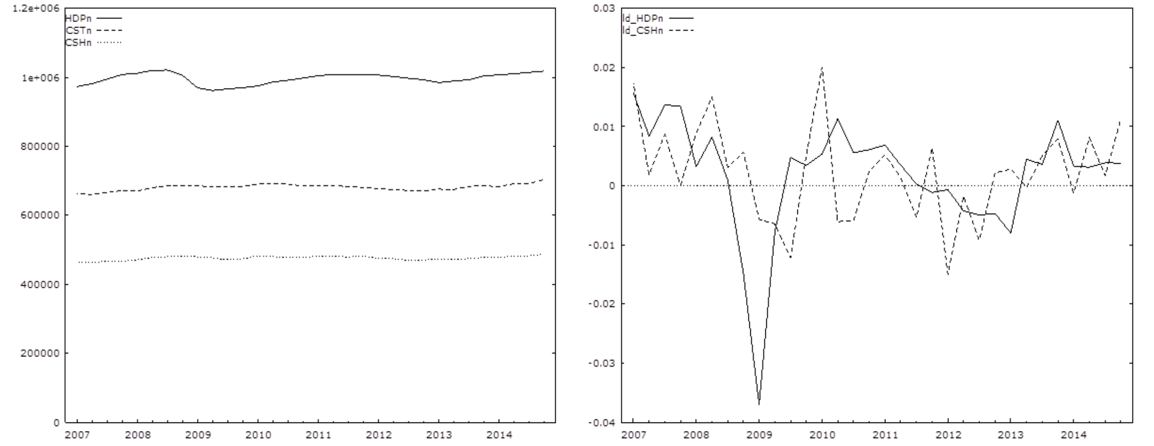
In the logit models for economic downturns, CCI remained the best among competing variables, with its fit measured by McFadden $R^2$ even increasing compared to original results, while the percentage of correct forecasts remained almost unchanged. Similar improvement in fit was apparent also for CLIs (but the objection described in the original thesis persists, CLIs would not be useful for nowcasting due to its publication delay), while the performance of PX worsened. Similarly to the original results, the CZK / EUR exchange rate proved the best fit in a longer horizon, with a peak on the 7$^{\text{th}}$ lag, with a fit improved compared with original results; but this may be slightly misleading since the exchange rate was held artificially at a relatively high level in 2014.

Most relevantly, GCSI also improved its fit and in accordance with the original results, its quality was persistent, as it belonged among the best even for distant lags, having both fit and the percentage of correct forecasts relatively high. Most importantly, GCSI proved superior to other variables when modeling the below average growth of household consumption. Its measure of fit improved and percentage of correct forecasts remained remarkably stable over several lags, predicting correctly 3 out of 4 crisis periods even 4 month ahead.

CCI also improved its fit significantly, together with a slight improvement of correct forecasts, but still remained inferior to GCSI, in accordance with original results for household consumption expenditures. PX index, Unemployment and BCI had worse results than before, while CLIs slightly improved. It should be noted that the prolonged period for logit models was relatively uneventful, as no crisis or below average growth period was added to the explained series; the general confirmation of original results is therefore not that surprising.

This is different for qualitative forecasts, as entire new series of GDP and household consumption were used for the period 2007-2014 (due to a change in methodology by the CZSO, as described above). For both of the variables, VAR(1) model was used for out-of-sample forecasting, with a recursively expanding window of an initial length $R = 20$ for parameters estimation and an out-of-sample period of length $P = 12$. The results are included in Appendix C.[36]

**Figure 8.2: Czech GDP, total consumption and household consumption (2007–2014)**




GDP (solid line), total consumption (dashed line), and household consumption (dotted line), constant prices of year 2010.

Quarterly growth rate of GDP (solid line) and household consumption (dashed line), constant prices of years 2010.

Source: CZSO, author's calculations

When comparing the forecasting accuracy on the out-of-sample period 2012-2014, GCSI performed the best among competing variables in non-nested models even for GDP. The striking difference compared with the original results is almost equal predictive accuracy of all variables, especially of GCSI, CLIs, BCI and CCI, and no definite conclusion can thus be made since no statistical significance was achieved. This was probably caused both by the substitution of the original GDP series by a new one as well as by the prolongation of the out-of-sample period.

On the other hand, for the household consumption forecasts, the results propose a better answer. GCSI performed better than other variables, statistically significant superior performance by more than 40% compared with the growth of Wage, CLIs, PX index and BCI. Even the improvement over CCI, which was not statistically significant, achieved 19%. These results reflect better the original results, as the order of variables remained the same (with the exception of PX index, whose deterioration is in line with the results of logit models).

Worse performance of all variables when forecasting the GDP growth was confirmed also in the trivariate VAR(1) setting, with a baseline model augmented by CLIs, where no statistically significant improvement was

[36] VAR(1) specification was chosen by BIC for most of the tested variables. Even for those where other specification was suggested, VAR(1) model provided the best forecasting accuracy.

achieved. In case of household consumption qualitative forecasts, the results are again in line with previously stated, GCSI performing the best and being followed by CCI. Most variables improved statistically significantly the baseline VAR(1) model with growth of average real wage. On the other hand, identically to the original results, one would be better off using a bivariate model for household consumption with GCSI (MSE of 59.6% relative to the bivariate VAR(1) model with Wage) instead of the best trivariate one (relative MSE 72.9%.

In overall, GCSI confirmed its qualities, especially for the household consumption, when extending the data set by one year. This is important as the original out-of-sample length was 8 observations, and the prolongation by 50% is rather significant to draw a better conclusion. We can thus see the performance of GCSI as quite stable. Forecasts of GDP were mostly affected by the fact that the series – over the analyzed period – was probably captured the best by its own lagged value, and additional explanatory variables could not do much more. In fact, a pure AR(1) process with a constant performed only slightly worse than the best GCSI model (the relative MSE was 105%, compared with 190% in a similar analysis for household consumption).

## 8.4 Stability of downloaded data

As described in Chapter 3 in the data description, the downloaded time series of Google data differ slightly depending on the day of download. For this reason, for the purpose of our analysis, we downloaded each time series for 28 consecutive days (in February 2014 for the original thesis and in April 2015 for results revision) and used the average time series as a representative for each query. An argument can thus be made against the potential of a practical use of Google data.

For this reason, we conducted an analysis to see if an easier construction was possible for real-life applications. For each of the Google queries analyzed in our thesis, we computed correlation coefficient between the overall 28-day average and other potential uses of Google data. First, this is the data from the first day of downloads, as if ignoring the effect of data variation depending on the day of download. Second, we calculated an average over first three days, and third, over first five days of data download. These versions of Google index would be indeed more practical than the 28-day average. The results for the analyzed period 2007-2014 are presented in Table (8.4).

**Table 8.4: Correlation coefficients of individual queries with the 28-day averages**

| | levels | | | log-differences | | |
|---|---|---|---|---|---|---|
| | 1st day | 3-day average | 5-day average | 1st day | 3-day average | 5-day average |
| **GCSI** | 0.975 | 0.987 | 0.987 | 0.873 | 0.923 | 0.920 |
| "job" | 0.996 | 0.998 | 0.995 | 0.949 | 0.970 | 0.933 |
| "job offers" | 0.989 | 0.996 | 0.999 | 0.710 | 0.825 | 0.982 |
| "crisis" | 0.994 | 0.989 | 0.989 | 0.933 | 0.819 | 0.393 |
| "furniture" | 0.986 | 0.996 | 0.998 | 0.843 | 0.949 | 0.972 |
| "inflation + price increase" | 0.845 | 0.916 | 0.917 | 0.729 | 0.741 | 0.749 |
| "distraint" | 0.976 | 0.997 | 0.999 | 0.642 | 0.920 | 0.956 |
| "saving + savings" | 0.960 | 0.974 | 0.987 | 0.833 | 0.934 | 0.917 |
| "employment" | 0.970 | 0.983 | 0.990 | 0.813 | 0.884 | 0.928 |
| "labor office" | 0.982 | 0.993 | 0.996 | 0.839 | 0.919 | 0.944 |
| "CV" | 0.997 | 0.998 | 0.999 | 0.897 | 0.925 | 0.968 |

Source: author's calculations

Explanation: The table displays correlation coefficients for the overall 28-day average of each Google query and individual series denoted in columns: (1) series from the first day of downloads, (2) average over first 3 days of downloads, (3) average over first five days of downloads. The left-hand side provides results for levels, the right-hand side for log-differences of each time series, calculated on seasonally adjusted data over a sample 2007-2014.

The left-hand side of the table provides results for levels of individual queries and GCSI (all seasonally adjusted), the right-hand side for log-differences. At first sight, it seems that downloading the data over 28 consecutive days was not necessary. For most of the queries, even data from the first day were well correlated with the average index. This improved (except for the case of ”crisis”) even more for 3 and 5-day averages, when the correlation coefficient was close to or above 0.99 for all queries except for ”inflation”.

Even more importantly, the same seems to hold also for log-differences, which is relevant especially for queries related to unemployment from our previous analysis. In this case, first day values have slightly lower correlation coefficients, but it improves steeply for the 3 and 5-day average (again with the exception of ”crisis”). Therefore, for the practical use of Google query data, we would recommend downloading each series more than only once – but a 3-day or 5-day average should suffice for the purposes of econometric analysis, especially for relatively popular web search queries.

# 9 Conclusion

Google Econometrics – the use of Google search volume data of particular queries as additional explanatory variables in time series modeling – has been tested so far mostly on developed economies, such as the United States, Canada, Japan, Germany, France, and other Western European countries. These methods were examined only sporadically for developing economies or emerging markets which differ not only in the characteristics of their economies, but also in the internet penetration rate; this translates both to the number of people using computers and the internet, but also to the level of internet skills and the frequency of use.

This thesis examined the applicability of Google Econometrics in the case of the Czech Republic, which has not been done up to the publication of the original thesis. On the data sample for years 2004–2013 (later updated to a period 2004–2014), we tested it in the following three areas: unemployment, consumer confidence, and overall economic situation. We conclude that Google Econometrics can be applied also to Czech data, and this is despite the fact that for most of the analyzed period the internet penetration rate was inferior to that in developed countries; and more importantly, Google does not dominate the Czech search engine market like it does in most of the world.

When modeling unemployment, search queries logically connected to the job searching process were used – "job", "job offers" or "Labor office" among others. We added these variables independently into a baseline autoregressive model and later into an autoregressive model with control variables, and tested the out-of-sample nowcasting performance of given queries. Google data improved nowcasts of the baseline models modestly (around 10% measured by Mean Squared Error) compared with related studies on other countries (where the improvement was sometimes even 50%). Nevertheless, this improvement was statistically significant and the best models contained Google data; so even in the case of the Czech Republic, Google data bring significant information not contained in control variables (such as Consumer Confidence, Index of Industrial Production or Composite Leading Indicators).

To assess consumers' sentiment, we combined two approaches employed in the related literature – we took the official Consumer Confidence Indicator as a benchmark and to each of the questions asked during official surveys, we assigned appropriate search queries. Based on the data of these queries, we created own Google Consumer Sentiment Index in a way analogous to the creation of the official indicator – normalizing and averaging individual time series. The resulting index correlated well with, captured well the dynamic of, and also appeared to lead the official indicator over the analyzed period. Granger causality test suggested that both indicators are closely interconnected. This is in accordance with related studies for other countries suggesting that people reveal valuable information about their sentiment by entering particular words into search engines.

The contribution of our index was tested on models for overall economic situation: the growth of GDP and household final consumption expenditures. We analyzed both in-sample fit when predicting economic downturns and below-average growth of consumption in the framework of logit models, and out-of-sample nowcasting accuracy using vector autoregression (VAR). For the original series of GDP growth, our Google index performed well, but did not reach the qualities of the best control variables, most notably the official Consumer Sentiment Indicator. On the other hand, it performed better than Business Confidence Indicator and only slightly worse than Composite Leading Indicators, which is designed to capture the development of GDP. This was true both for logit and VAR models.

When modeling household consumption, Google Consumer Sentiment Index was the best among the set of control variables, including the official Consumer Confidence Indicator. This was confirmed both by in-sample fit of logit models for predicting probability of below-average growth, as well as by out-of-sample nowcasting accuracy where Google index also significantly improved the quality of models containing control variables. This is in accordance with the results of studies conducted on consumption in other countries – people reveal information about their potential purchasing plans (or their economic situation or sentiment that translates into their consumption expenditures) by searching for particular queries.

To conclude, we found that Google data contain valuable information even in the case of the Czech Republic. This was confirmed also on the dataset extended by one year, as well as by other two studies conducted on this topic for the Czech Republic. Because the performance of Google data was better for nowcasting, this data may be analyzed to provide information about the current situation of economic agents, in addition to leading and coincident indicators usually used. The most notable difference of the Czech Republic to compared countries was that Google does not dominate the search engine market, which might be a concern when using volumes of searches. However, in accordance to this and also to the lower internet penetration rate in the beginning of the analyzed period, Google data contained some noise; also because many queries were not searched for enough in absolute terms, their time series were incomplete. For this reason, we suggest using this data with a restricted sample, for example from 2007 onwards like in our analysis of sentiment and overall economic situation.

In addition to the topics explored in this thesis, there are a lot of other areas where Google data may be useful. In the related literature, many studies concentrated on the interconnection of search volumes and financial markets, using it either as a proxy for investors' sentiment or their attention to particular stocks. Often, the authors found that such data have good explanatory power, and some showed that trading strategy based on this data is profitable; this can be also tested for Czech stocks traded at the Prague Stock Exchange.

In our thesis, consumption was analyzed as an aggregate macroeconomic variable, but a detailed look at particular queries may also reveal more specific information. For example, the interconnection between the volume of searches of car brands and their sales may be examined, similarly for other types of goods. In such cases, however, an attention has to be paid to the completeness of the data of particular queries – this is not a problem for the use in the future, but rather for backwards testing of an appropriate choice of a query.

For our own work, we would like to extend the study of our Google Consumer Sentiment Index in a way that would copy the real-time creation of sentiment indicators; that means a recursive scheme (using all available data) to obtain the value of the index for one month at a time. This approach could not be employed in this thesis given a short data sample of the queries used; but this can only get better given the increased internet penetration rate as well as Google search engine popularity in the Czech Republic in recent years. The index created based on data prolonged by an additional year showed promising results, as its interconnection with the official indicator persisted.

Social media feeds, such as Twitter, represent another possible extension of the analysis of internet data application on the economy of the Czech Republic. This has been tested already extensively for the U.S. data, but an application to other countries (and other languages) is still relatively scarce. Given the popularity of Twitter in the Czech Republic and also the characteristics of the Czech language, conducting an analysis analogous to those described in the literature review proposes a challenging task.

# Bibliography

**Antenucci, D., Cafarella, M., Levenstein, M.C., Ré, C., Shapiro, M.D. (2014)**: Using Social Media to Measure Labor Market Flows, NBER Working Paper No. 20010, March 2014

**Antweiler, W., Frank, M. (2001)**: Is All The Talk Just Noise? The Information Content of Internet Stock Message Board, *The Journal of Finance*, Vol. 59, No. 3, pp. 1259-1294, June 2004

**Askitas, N., Zimmermann, K. (2009)**: Google Econometrics and Unemployment Forecasting, IZA Discussion Paper No. 4201, June 2009

**Asur, S., Huberman, B. (2010)**: Predicting the Future with Social Media, arXiv.org, Computers and Society Papers 1003.5699, March 2010

**Bańbura, M., Giannone, D., Reichlin, L. (2010)**: Nowcasting, European Central Bank, Working Paper No. 1275, December 2010

**Bank, M., Larch, M., Peter, G. (2011)**: Google Search Volume and its Influence on Liquidity and Returns of German Stocks, *Financial Markets and Portfolio Management*, Vol. 25, No. 3, pp. 239-264, 2011

**Beber, A., Brandt, M.W., Luisi, M. (2013)**: Distilling the Macroeconomic News Flow, Center for Economic Policy Research, Discussion Paper No. 9360, February 2013

**Becchetti, L., Marini, G., Murgea, A. (2012)**: The Spread Nightmare: Financial Crisis and Happiness, Tor Vergata University, CEIS Research Paper series, Vol. 10, Issue 12, No. 252, September 2012

**Beer, F., Hervé, F., Zouaoui, M. (2013)**: Is Big Brother Watching Us? Google, Investor Sentiment and the Stock Market, AccessEcon, *Economics Bulletin*, Vol. 33, Issue 1, pp. 454-466, 2013

**Bollen, J., Mao, H., Zeng, X.J. (2011)**: Twitter Mood Predicts the Stock Market, *Journal of Computational Science*, Vol. 2, Issue 1, pp. 1-8, March 2011

**Bollen, J., Pepe, A., Mao, H. (2009)**: Modeling Public Mood and Emotion: Twitter Sentiment and Socioeconomic Phenomena, arXiv.org, Computers and Society Papers 0911.1583, November 2009

**de Bondt, G.J., Schiaffi, S. (2011)**: Confidence Matters for Nowcasting GDP: Euro Area and US Evidence from a PMI-based Model, European Central Bank, Working Paper Series, November 2011

**Brooks, C. (2008)**: Introductory Econometrics for Finance, 2nd edition, Cambridge University Press, ISBN-13 978-0-521-69468-1, 2008

**Carrière-Swallow, Y., Labbé, F. (2010)**: Nowcasting with Google Trends in an Emerging Market, Central Bank of Chile Working Paper Nr. 588, July 2010

**Chadwick, M.G., Sengul, G. (2012)**: Nowcasting Unemployment Rate in Turkey: Let's Ask Google, Central Bank of the Republic of Turkey, Monetary Policy Department, Working Paper No. 1218, June 2012

**Challet, D., Ayed, A. (2014)**: Do Google Trends Data Contain More Predictability Than Price Returns?, arXiv.org Papers 1403.1715, March 2014

**Chamberlin, G. (2010)**: Googling the Present, UK Office for National Statistics, *Economic & Labour Market Review*, pp. 59-95, December 2010

**Choi, H., Varian, H. (2012)**: Predicting the Present with Google Trends, The Economic Society of Australia, *The Economic Record*, Vol. 88, Special Issue, pp. 2-9, June 2012

**Choi, H., Varian, H. (2009)**: Predicting the Present with Google Trends, Google Inc., April 2009

**Clark, T.E., McCracken, M.W. (2011)**: Advances in Forecast Evaluation, Federal Reserve Bank of Cleveland, Working Paper 11-20, September 2011

**Clark, T.E., West, K.D. (2007)**: Approximately Normal Tests for Equal Predictive Accuracy in Nested Models, *Journal of Econometrics*, Elsevier, vol. 138(1), pp. 291-311, May 2007

**Culotta, A. (2010)**: Towards Detecting Influenza Epidemics by Analyzing Twitter Messages, Proceedings of the First Workshop on Social Media Analytics (Soma '10), pp. 115-122, July 2010

**Curini, L., Iacus, S., Canova, L. (2015)**: Measuring Idiosyncratic Happiness Through the Analysis of Twitter: An Application to the Italian Case, *Social Indicators Research*, Vol. 121, Issue 2, pp. 525-542, April 2015

**Da, Z., Engelberg, J., Gao, P. (2013)**: The Sum of All FEARS: Investor Sentiment and Asset Prices, Working paper for The Journal of Finance, 2013

**Da, Z., Engelberg, J., Gao, P. (2011)**: In Search of Attention, *The Journal of Finance*, Vol. 66, No. 5, pp. 1461-1499, October 2011

**D'Amuri, F., Marcucci, J. (2012)**: The Predictive Power of Google Searches in Forecasting Unemployment, Bank of Italy, Economic Research and International Relations Area, Working Paper No. 891, November 2012

**D'Amuri, F., Marcucci, J. (2010)**: "Google it!" Forecasting the US Unemployment Rate with a Google Job Search Index, FEEM Working Paper No. 31, 2010

**D'Amuri, F. (2009)**: Predicting Unemployment in Short Samples with Internet Job Search Query Data, MPRA Paper No. 18403, October 2009

**Das, P.K., Ziobrowski, A.J. (2015)**: The Relationship between Indian Realty Stocks and Online Searches, *Journal of Emerging Market Finance*, Vol. 14 No. 1, pp. 1-19, April 2015

**Dées, S., Brinca, P.S. (2011)**: Consumer Confidence as a Predictor of Consumption Spending: Evidence for the United States and the Euro Area, European Central Bank, Working paper series, No. 1349 / June 2011

**Della Penna, N., Huang, H. (2009)**: Constructing Consumer Sentiment Index for U.S. Using Internet Search Patterns, University of Alberta, Department of Economics, Working Paper No. 2009-26, 2009

**Dergiades, T., Milas, C., Panagiotidis, T. (2013)**: Tweets, Google Trends and Sovereign Spreads in the GIIPS, *Oxford Economic Papers*, Oxford University Press, Vol. 67 Issue 2, pages 406-432, 2013

**DG-ECFIN (2007)**: The Joint Harmonised EU Programme of Business and Consumer Survey, Directorate General for Economic and Financial Affairs, 2007

**Diebold, F.X. (2012)**: Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests, PIER Working Paper 12-035, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania, 2012

**Diebold, F.X., Mariano, R.S. (1995)**: Comparing Predictive Accuracy, *Journal of Business & Economic Statistics*, American Statistical Association, vol.13 (3), pp. 253-64, July 1995

**Ding, R., Hou, W. (2011)**: Retail Investor Attention and Stock Liquidity, eBusiness & eCommerece eJournal 03/2011, 2011

**Döhrn, R. (2013)**: Transportation Data as a Tool for Nowcasting Economic Activity: The German Road Pricing System as an Example, RUHR Economic Paper No. 395, January 2013

**Dzielinski, M. (2012)**: Measuring Economic Uncertainty and its Impact on the Stock Market, *Financial Research Letters*, Vol. 9, Issue 3, pp. 167-175, 2012

**Earle, P., Guy, M., Buckmaster, R., Ostrum, C., Horvath, S., Vaughan, A. (2010)**: OMG Earthquake! Can Twitter Improve Earthquake Response?, *Seismological Research Letters*, Vol. 81, Nr. 2, March/April 2010

**Enders, W. (2003)**: Applied Econometric Time Series, 2nd edition, Wiley, ISBN 978-0-471230656, 2003

**Estrella, A., Mishkin, F.S. (1998)**: Predicting U.S. Recessions: Financial Variables As Leading Indicators, *The Review of Economics and Statistics*, MIT Press, Vol. 80, Issue 1, pp 45-61, February 1998

**Fišer, R. (2010)**: Confidence, Economy and the Global Economic Crisis, master thesis, Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies, 2010

**Fondeur, Y., Karamé, F. (2013)**: Can Google Data Help Predict French Youth Unemployment?, *Economic Modeling*, Vol. 30, pp. 117-125, January 2013

**Gerow, A., Keane, M. (2011)**: Mining the Web for the 'Voice of Herd' to Track Stock Market Bubbles, Proceedings of the International Joint Conference on Artificial Intelligence, July 2011

**Giannone, D., Reichlin, L., Simonelli, S. (2009)**: Nowcasting Euro Area Economic Activity in Real-Time: The Role of Confidence Indicators, ECARES Working Paper No. 2009-021, September 2009

**Gilbert, E., Karahalios, K. (2010)**: Widespread Worry and the Stock Market, Proceedings of the AAAI Conference on Weblogs and Social Media, 2010

**Gill, T., Percra, D., Sunner, D. (2012)**: Electronic Indicators of Economic Activity, Reserve Bank of Australia Bulletin, June Quarter 2012

**Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L. (2009)**: Detecting Influenza Epidemics Using Search Engine Query Data, *Nature*, Vol. 457, 19 February 2009

**Golder, S.A., Macy, M.W. (2011)**: Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures, *Science*, Vol. 333, pp. 1878-1881, September 2011

**Gómez, V., Maravall, A. (1997)**: TRAMO and SEATS: Instructions for the User, Bank of Spain, November 1997

**Greene, W.H. (2011)**: Econometric Analysis, 12th edition, Prentice Hall, ISBN-13 978-0-13139538-1, 2011

**Guan, X., Chen, C. (2014)**: Using Social Media Data to Understand and Assess Disasters, *Natural Hazards*, Vol. 74, Issue 2, pp. 837-850, November 2014

**Hansen, P.R., Timmermann, A. (2012)**: Choice of Sample Split in Out-of-Sample Forecast Evaluation, CREATES Research Papers 2012-43, February 2012

**Harvey, D.I., Leybourne, S.J., Newbold, P. (1997)**: Testing the Equality of Prediction Mean Squared Errors, Elsevier, *International Journal of Forecasting*, vol. 13, pp. 281-291, June 1997

**Herrmannová, L. (2012)**: Forecasting Ability of Confidence Indicators: Evidence for the Czech Republic, master thesis, Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies, 2012

**Hohenstatt, R., Käsbauer, M., Schäfers, W. (2011)**: "Geco" and its Potential for Real Estate Research: Evidence from the US Housing Market, American Real Estate Society, *Journal of Real Estate Research*, Vol. 33, Issue 4, pp. 471-506, 2011

**Horváth, R. (2012)**: Do Confidence Indicators Help Predict Economic Activity? The Case of the Czech Republic, *Czech Journal of Economic and Finance*, Vol. 62 (2012), Issue 5 (November), pp. 398-412, 2012

**Iselin, D., Siliverstovs, B. (2013)**: Using Newspapers for Tracking the Business Cycle: A Comparative Study for Germany and Switzerland, KOF Swiss Economic Institute, Working Paper No. 337, June 2013

**Janetzko, D. (2014)**: Predictive Modeling in Turbulent Times – What Twitter Reveals about the EUR/USD Exchange Rate, *NETNOMICS: Economic Research and Electronic Networking*, Vol. 15, Issue 2, pp. 69-106, September 2014

**Kholodilin, K.A., Podstawski, M., Siliverstovs, B. (2010)**: Do Google Searches Help in Nowcasting Private Consumption? A Real-Time Evidence for the US, KOF Working Papers No. 256, April 2010

**Ko, J., Kwon, H.W., Kim, H.S., Lee, K., Choi, M.Y. (2014)**: Model for Twitter Dynamics: Public Attention and Time Series of Tweeting, *Physica A: Statistical Mechanics and its Applications*, Vol. 404, pp. 142-149, June 2014

**Krištoufek, L. (2013)**: Can Google Trends Search Queries Contribute to Risk Diversification?, Scientific Reports Papers 3:2713, October 2013

**Kulkarni, R., Haynes, K.E., Stough, R.R., Paelinck, J.H.P. (2009)**: Forecasting Housing Prices with Google Econometrics, George Mason University, School of Public Policy, Research Paper No. 2009-10, 2009

**Ludvigson, S.C. (2004)**: Consumer Confidence and Consumer Spending, *Journal of Economic Perspectives*, Vol. 18, Nr. 2, pp. 29-50, Spring 2004

**Maddala, G.S. (2001)**: Introduction to Econometrics, 3rd edition, Wiley, ISBN 978-0-471497288, 2001

**Mao, H., Counts, S., Bollen, J. (2011)**: Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data, arXiv.org, Quantitative Finance Papers 1112.1051, December 2011

**OECD (2014)**: OECD Composite Leading Indicators: Turning Points of Reference Series and Component Series, April 2014

**OECD (2012)**: OECD System of Composite Leading Indicators, April 2012

**Papaioannou, P., Russo, L., Papaioannou, G., Siettos, C.I. (2013)**: Can Social Microblogging Be Used to Forecast Intraday Exchange rates?, *Journal Netnomics*, Vol. 14, Issue 1-2, pp. 47-68, November 2013

**Pavlíček, J., Krištoufek, L. (2015)**: Nowcasting Unemployment Rates with Google Searches: Evidence from the Visegrad Group Countries, FinMaP-Working Paper No. 34, March 2015

**Philippas, D., Siriopoulos, C. (2012)**: The Effect of Time-Varying Value of Information on Market Uncertainty, Joint Research Center of the European Commission Working Paper, May 2012

**Porshnev, A., Redkin, I., Shevchenko, A. (2013)**: Improving Prediction of Stock Market Indices by Analyzing the Psychological States of Twitter Users, Higher School of Economics Research Paper No. WP BRP 22/FE/2013

**Preis, T., Reith, D., Stanley, H. (2010)**: Complex Dynamics of Our Economic Life on Different Scales: Insights from Search Engine Query Data, *Philosophical Transactions of the Royal Society*, vol. 368, issue 1933, pp. 5707-5719, 2010

**Ripberger, J. (2011)**: Capturing Curiosity: Using Internet Search Trends to Measure Public Attentiveness, *Policy Studies Journal*, Vol. 39, Issue 2, pp. 239-259, May 2011

**Rubin, A., Rubin, E. (2010)**: Informed Investors and the Internet, *Journal of Business Finance & Accounting*, Vol. 37, Issue 7-8, pp. 841-865, July / August 2010

**Saxa, B. (2014)**: Forecasting Mortgages: Internet Search Data as a Proxy for Mortgage Credit Demand, Czech National Bank Working Paper Series 14/2014, December 2014

**Schmidt, T., Vosen, S. (2012)**: A Monthly Consumption Indicator for Germany Based on Internet Search Query Data, *Applied Economics Letters*, Vol. 19, Issue 7, pp. 683-687, May 2012

**Schmidt, T., Vosen, S. (2011)**: Forecasting Private Consumption: Survey-based Indicators vs. Google Trends, *Journal of Forecasting*, Vol. 30, Issue 6, pp. 565-578, September 2011

**Sims, C.A. (1980)**: Macroeconomics and Reality, *Econometrica*, Vol. 4, January 1980, pp. 1-48

**Souza, T.T.P., Kolchyna, O., Treleaven, P.C., Aste, T. (2015)**: Twitter Sentiment Analysis Applied to Finance: A Case Study in the Retail Industry, arXiv.org, Computers and Society Papers 1507.00784, July 2015

**Stephens-Davidowitz, S. (2013)**: The Cost of Racial Animus on a Black Presidential Candidate: Using Google Search Data to Find What Surveys Miss, Harvard University - Department of Economics, Working Paper series, March 2013

**Stock, J.H., Watson, M.W. (2007)**: Introduction to Econometrics, 2nd edition, Pearson/Addison Wesley, ISBN 978-0-321442536, 2007

**Su, Z. (2014)**: Chinese Online Unemployment-Related Searches and Macroeconomic Indicators, *Frontiers of Economics in China*, Vol. 9, Issue 4, pp. 573-605, December 2014

**Suhoy, T. (2009)**: Query Indices and a 2008 Downturn: Israeli Data, Bank of Israel, Research Department, Discussion Paper No. 2009.06, July 2009

**Takeda, F., Wakao, T. (2014)**: Google Search Intensity and Its Relationship with Returns and Trading Volume of Japanese Stocks, *Pacific-Basin Finance Journal*, Vol. 27, pp. 1-18, April 2014

**Taylor, K., McNabb, R. (2007)**: Business Cycles and the Role of Confidence: Evidence for Europe, *Oxford Bulletin of Economics and Statistics*, Department of Economics, University of Oxford, Vol. 69, Issue 2, pp. 185–208,2007

**Tetlock, P. (2007)**: Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *The Journal of Finance*, Vol. 62, No. 3, pp. 1139-1168, June 2007

**Tkacz, G. (2013)**: Predicting Recessions in Real-Time: Mining Google Trends and Electronic Payments Data for Clues, C.D. Howe Institute Commentary, No. 387, September 2013

**Tuhkuri, J. (2014)**: Big Data: Google Searches Predict Unemployment in Finland, ETLA (The Research Institute of the Finnish Economy), Reports No. 31, August 2014

**Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M. (2010)**: Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment, Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, May 2010

**Uhl, M.W. (2011)**: Nowcasting Private Consumption with TV Sentiment, KOF Swiss Economic Institute, Working Paper No. 293, December 2011

**Varian, H.R. (2014)**: Big Data: New Tricks for Econometrics, *Journal of Economic Perspectives*, Vol. 28, Nr. 2, pp. 3-28, Spring 2014

**Vlastakis, N., Merkellos, R. (2012)**: Information Demand and Stock Market Volatility, *Journal of Banking and Finance*, Vol. 36, No. 6, pp. 1808-1821, 2012

**West, K.D. (2006)**: Forecast Evaluation, Handbook of Economic Forecasting, Elsevier, Vol. 1, 2006

**Wu, L., Brynjolfsson, E. (2013)**: The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales, MIT Sloan School of Management, Research Paper, 2013

**Yang, A.C., Huang, N.E., Peng, C.-K., Tsai, S.-J. (2010)**: Do Seasons Have an Influence on the Incidence of Depression? The Use of an Internet Search Engine Query Data as a Proxy of Human Affect, PLoS One, Vol. 5, Issue 10, e13728, October 2010

**Zhang, X., Fuehres, H., Gloor, P.A. (2011)**: Predicting Stock Market Indicators Through Twitter 'I hope it is not as bad as I fear', *Procedia – Social and Behavioral Sciences*, Vol. 26, pp. 55-62, 2011

**Electronic sources:**

| | |
|---|---|
| www.google.com/trends | Google Trends |
| www.czso.cz | Czech Statistical Office |
| stats.oecd.org | OECD |
| www.cnb.cz/docs/ARADY/HTML/index.htm | Czech National Bank |
| epp.eurostat.ec.europa.eu | Eurostat |
| www.pse.cz | Prague Stock Exchange |
| www.comscore.com | comScore.com |
| www.atinternet.com | AT Internet |
| www.navrcholu.cz | Navrcholu.cz |
| www.toplist.cz | TOPlist |
| www.oxforddictionaries.com | Oxford Dictionaries |
| www.businessinsider.com | Business Insider |
| www.ccs.neu.edu | Northeastern University |
| www.bluenileresearch.com | Blue Nile Research |

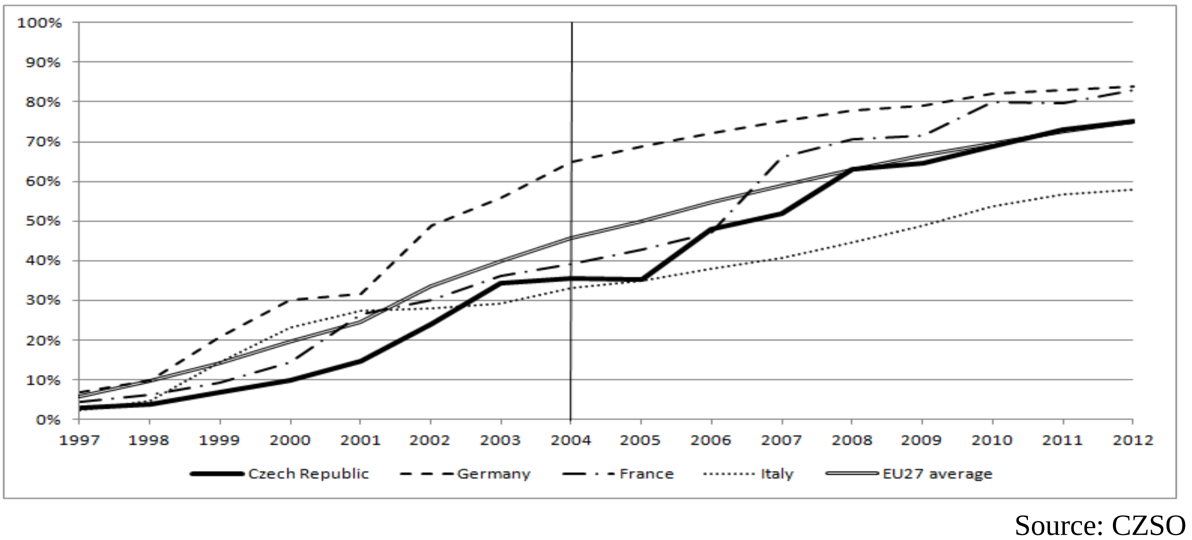
# Appendix A: Internet penetration rate, Internet skills, Google popularity

When we analyze the connection between search query data and real economic activity, it is beneficial to understand what proportion of population uses the internet, what are the skills of internet users, and last but not least, how much they use Google when looking for information through search engine. This Appendix provides a brief statistical description of this; the source of statistics presented here are the Czech Statistical Office (CZSO) and Eurostat. [37]

## A.1 Internet penetration

Figure (A.1) shows the development of the share of internet users in the Czech Republic and few European countries (where Google Econometrics has been applied) over the period 1997–2012. Vertical line in year 2004 denotes the availability of Google data. In 1997, internet was uncommon among European population, but the shares started to rise soon after that. In 2004, it was 65% in Germany, EU27[38] average was 46%. The Czech Republic was below the average with 35%, behind France but slightly in front of Italy.

**Figure A.1: Share of internet users – European countries (population 16–74 years old)**



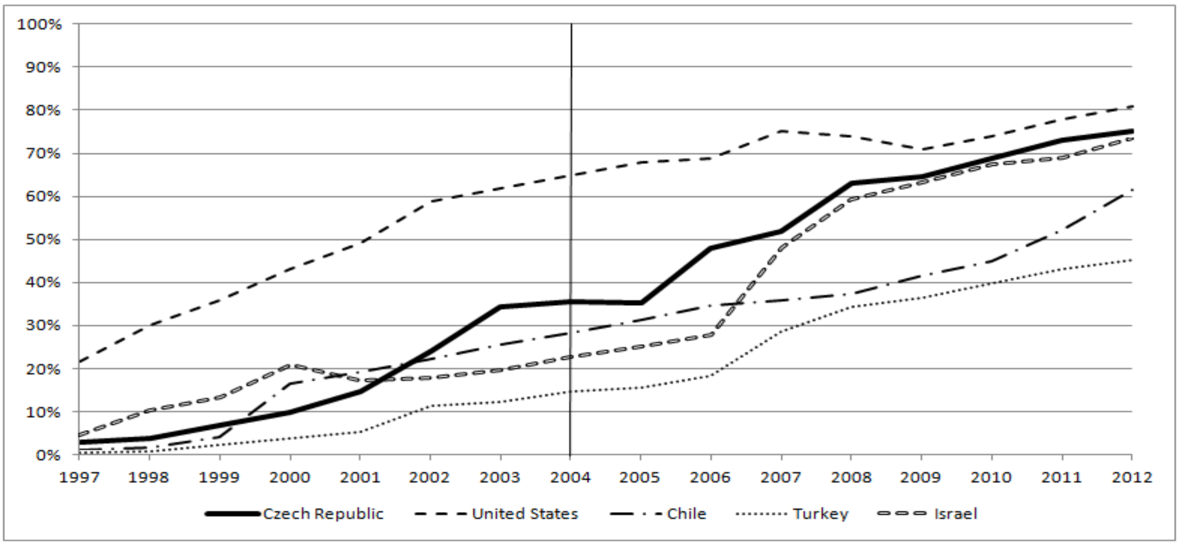

Source: CZSO

[37] Their methodology slightly differs: CZSO provides data about population of age 16+, Eurostat of age 16–74. For this reason, figures directly comparing different countries provide values by Eurostat (16–74), description of Czech characteristics follows the approach of CZSO (16+).

[38] EU27 denotes the group of European Union member countries in the period 2007–2013; even though it did not exist before 2007, Eurostat and CZSO provide the data also for preceding years.

The Czech Republic reached the average of EU27 by 2008 and remained at it, with 75% in 2012; France was still above the Czech share and Germany well above; on the other hand, Italy was constantly behind with less than 60% even in 2012. Figure (A.2) depicts the same data for some other countries to which Google Econometrics was also applied.

**Figure A.2: Share of internet users – various countries (population 16–74 years old)**

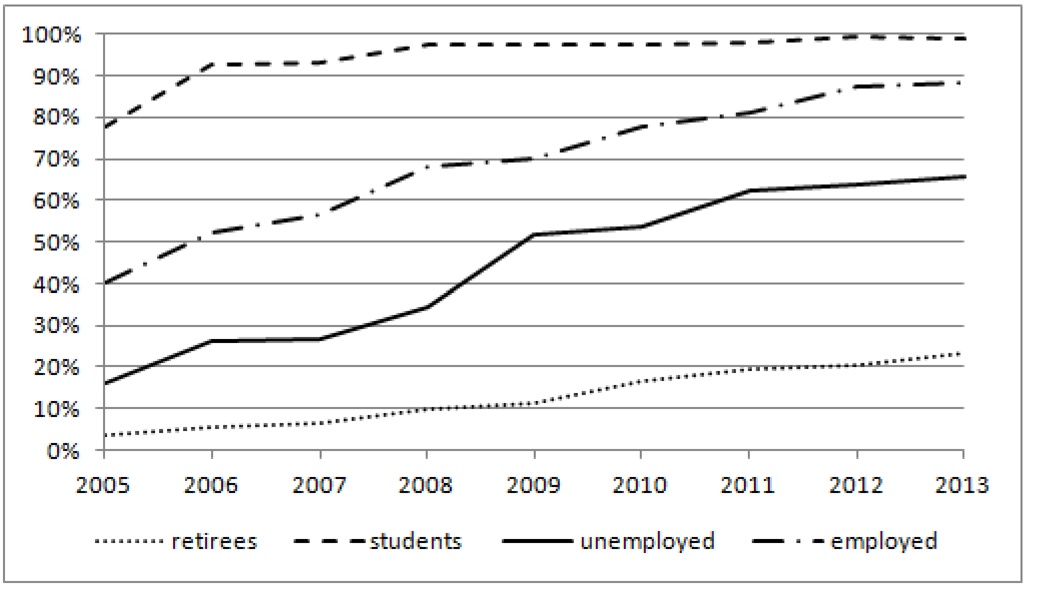



Source: CZSO

The starting position of the United States was superior to previously analyzed countries in 1997, but in 2004, the share was at 65%, the same as in Germany; the Czech Republic was at half of the value at the time, but it kept closing the gap. Comparing the Czech Republic to other emerging market, the share of internet users was similar to Israel over the analyzed period; Chile and Turkey were relatively far behind. In conclusion, Google Econometrics has been successfully applied to countries with a lower share of internet users than in the Czech Republic – not only emerging markets, but also Italy.

Within countries, the share of internet users differs also between individual groups – for example by age (gradually from youngest to oldest) or by education (from highest to lowest attained education) – this is true both for the Czech Republic as well as for the compared European countries. Figure (A.3) shows the division by socio-economic status in the Czech Republic, depicting obvious difference between employed and unemployed people.

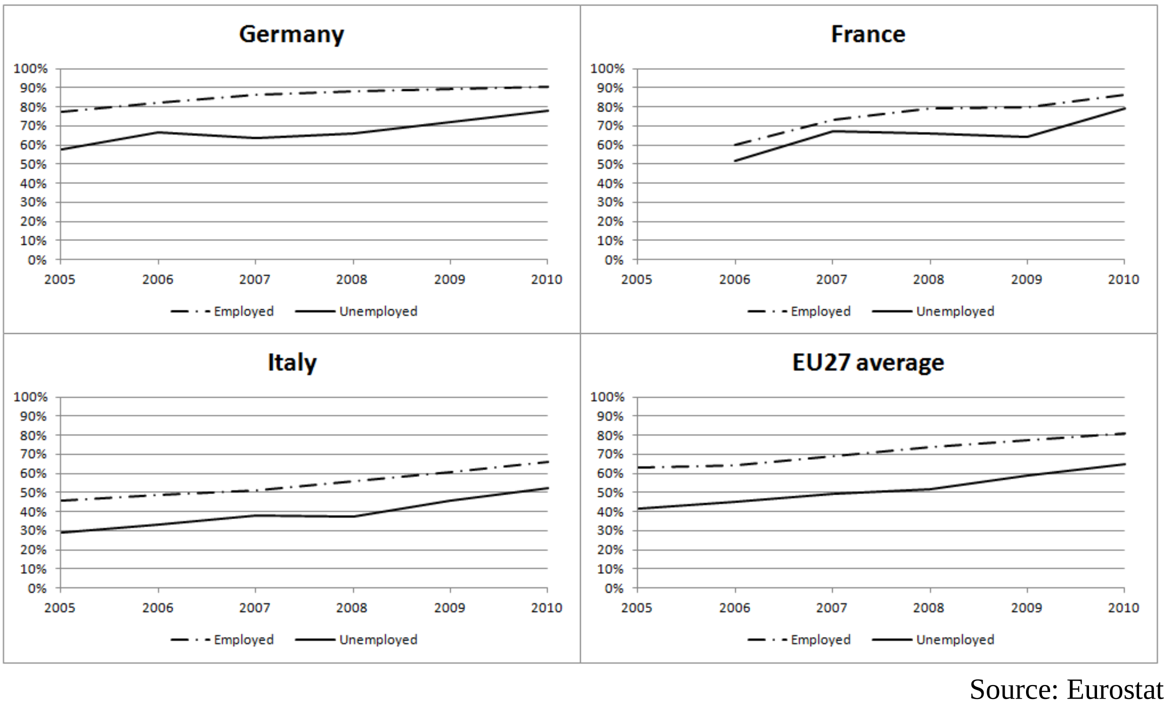
**Figure A.3: Share of internet users in groups by socio-economic status – Czech Republic – population 16+ (2005–2013)**




Source: CZSO

This difference is not exclusive to the Czech Republic, Figure (A.4) describes the same for the compared European countries. On the other hand, the actual share of unemployed using the internet was lower in the Czech Republic compared with these countries, as it attained the EU27 average only gradually, still lagging 10 p.p. in 2010 – far behind Germany and France, but in front of Italy.

**Figure A.4: Share of internet users among employed and unemployed – Europe – population 16–74 (2005–2010)**

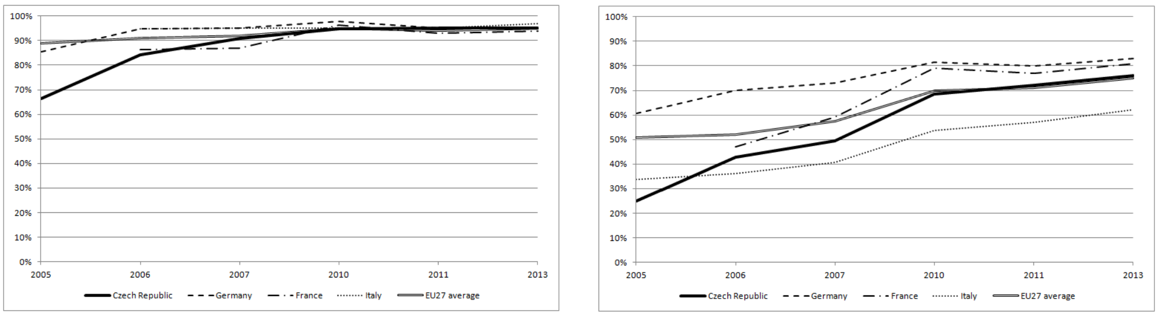



Source: Eurostat

## A.2 Internet skills and activities

Eurostat database defines six categories of internet skills, one of them being ”using a search engine to find information”. Figure (A.5) shows the shares of internet users and population over the period 2005–2013 in the compared European countries.[39] Over the analyzed period, most of the users possessed this skill – only in the Czech Republic in 2004, it was 66%, but it increased to the average of EU27 by 2007 (LHS chart).

**Figure A.5: Internet skills – using a search engine to find information (2005–2013)**

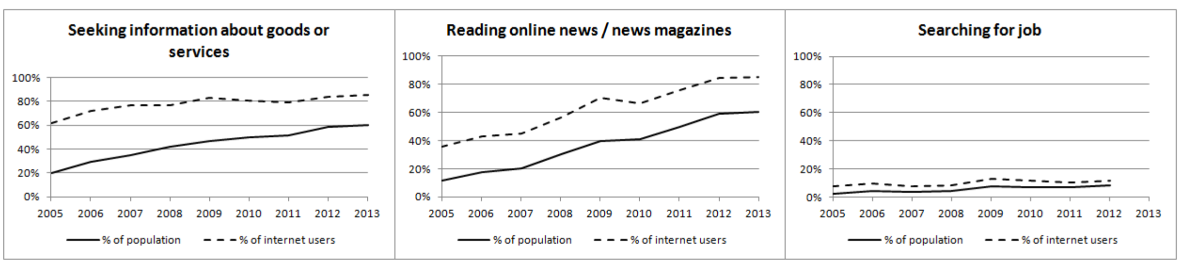



Share of internet users, 16-74. Share of population, 16-74.

Source: Eurostat

When internet penetration is taken into account – and the values are analyzed as a share of the whole population – the Czech Republic was well below average in the beginning of the analyzed period, 25% in 2005 compared with 50% for EU27, but it got closer to France and above Italy the next year and started to approach the EU27 average. Figure (A.6) shows the popularity of selected activities among Czech internet users (and also as a percentage of population).[40]

**Figure A.6: Selected internet activities – Czech Republic – 16+ (2005–2013)**



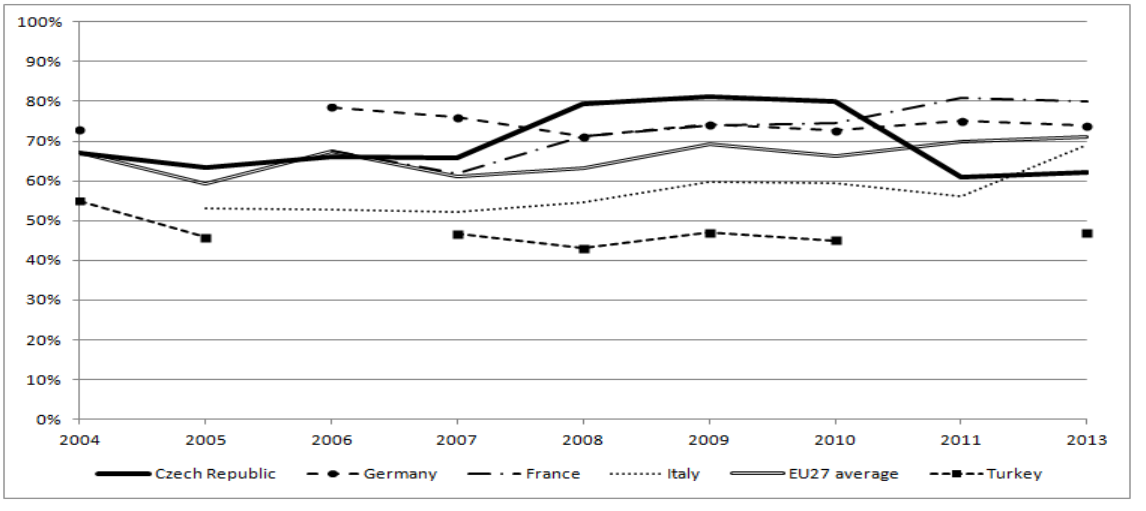

Source: CZSO

[39] In the LHS chart, it is the share of people with this skill to all people who ever used the internet. Also, please note that the time series is not complete; values for years 2008, 2009 and 2012 were not available.

[40] The share accounts for those who conducted this activity in three months preceding the survey.

Seeking information about goods and services is popular among Czech internet users, as well as reading online news and news magazines. Compared with that, searching for job was not that popular; this can be explained by the fact that unlike the aforementioned activities, searching for job is relevant only for a specific part of the population. Figure (A.7) depicts this share among unemployed internet users (who used the internet in last 3 months) in selected European countries.

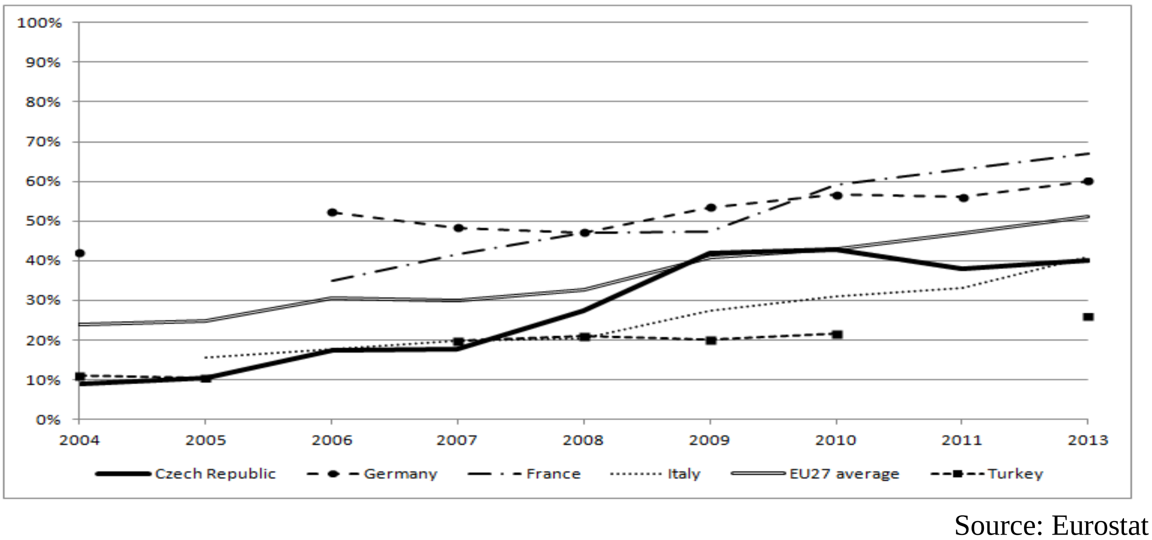
**Figure A.7: Share of unemployed internet users searching for job on the internet**




Source: Eurostat

From this point of view, the Czech Republic was an average country over past 10 years, having the share at least 60%; for example in Turkey, this share was below 50%. But as mentioned above, countries differ in the internet penetration rate among unemployed, Figure (A.8) shows the development when this is taken into account, showing the share of all unemployed in the population who searched for job on the internet.

**Figure A.8: Share of unemployed population searching for job on the internet**




Source: Eurostat

This picture is very different from the previous figure. It was only 9% in the Czech Republic in 2004; Germany had 42% already and the EU27 average was 24%. Over the whole period, the share of unemployed searching for job over the internet in the Czech Republic was inferior to that in France and Germany, but it equalized the EU27 average in 2009 and surpassed both Italy and Turkey by 2008 – in both of these countries, Google Econometrics in the area of unemployment was successfully applied.

## A.3 Google popularity

Google has been the most popular search engine around the world for some time already (measured as a percentage of searches conducted through Google search engine to the total number of searches). As a consequence, its name has become and English verb – a synonym for this action – recognized by official dictionaries; for example by American Merriam-Webster, who started to include this word in 2006, or British Oxford English Dictionary, who defines:

> google: search for information about (someone or something) on the Internet using the search engine.[41]

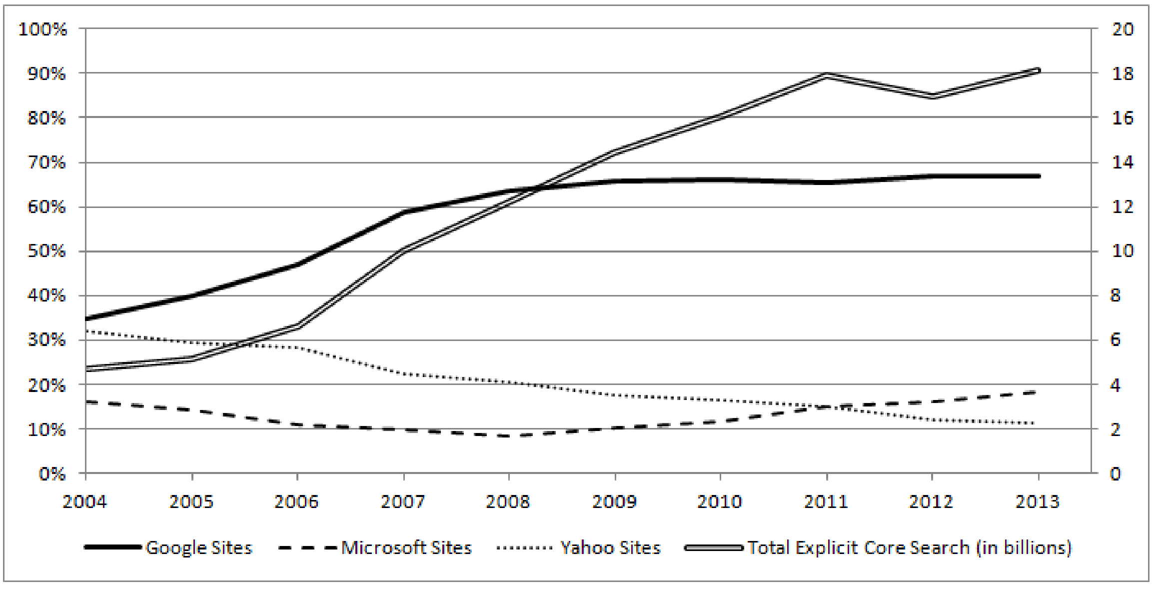
In spoken language, similar tendencies appeared also in other languages, such as in French (”googler”) or even in Czech, but it is not officially recognized yet.

Unlike the case of internet penetration and internet activities, there is no official source of data about search engines popularity around the world; the figures are provided by private companies who offer internet marketing services to their customers. Measuring shares of search engines helps them in optimizing their services and public releases of the data are rather meant to attract customers.

Figure (A.9) describes the popularity of Google in the United States – the state of both the origin of the company as well as of Google Econometrics. Even there, its position was not dominant in 2004; the share was 35%, almost equal to Yahoo. But by 2007, Google reached 60% and by the end of the period, 2 of 3 searches in the United States were conducted through Google. The RHS scale measures the total number of searches in November each year (doubled line), showing that it grew almost four times over the period, so the increase in absolute volume of searches through Google rose 7.4 times.

[41] http://www.oxforddictionaries.com/definition/english/google

**Figure A.9: Share of search engines in the United States (2004–2013)**

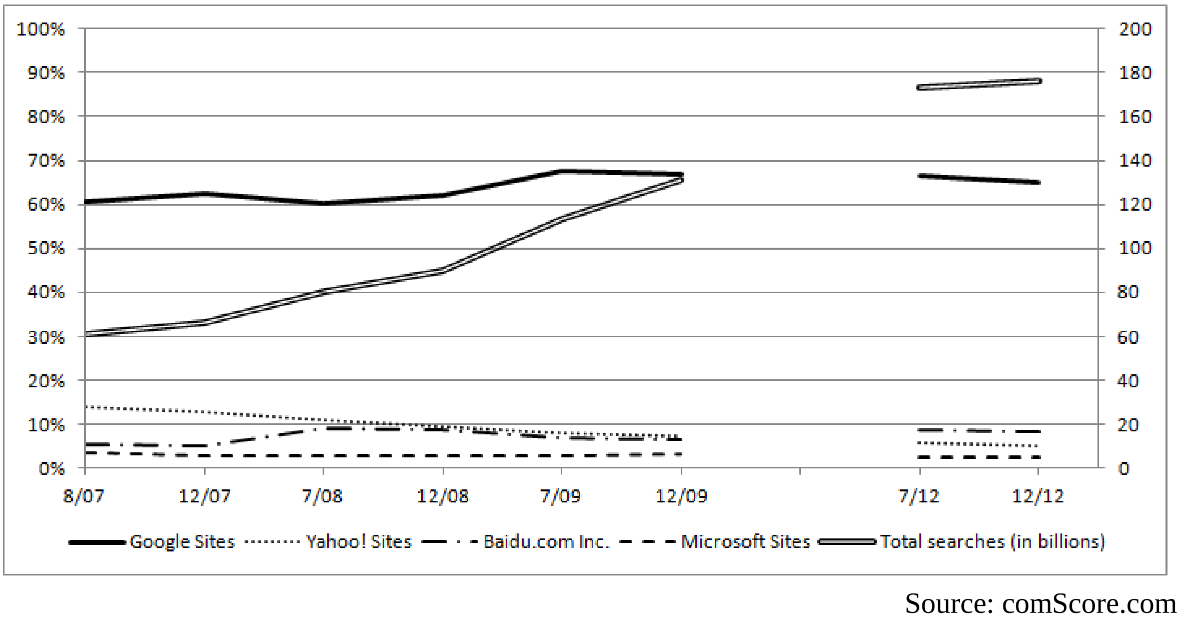



Source: comScore.com

Explanation: November value of each year is assigned to that year – this is relevant especially to the total number of searches conducted (values in the chart are not yearly).

Figure (A.10) describes the same information worldwide, where the share of Google was more than 60% in 2004 already, and grew slightly to similar values like in the U.S. until 2013; all other engines were below 10% for most of the time. The total number of searches increased almost three times over the period.

**Figure A.10: Share of search engines worldwide (2004–2013)**




Source: comScore.com

Explanation: Values for a given month are displayed, this is relevant especially to the total number of searches in the RHS scale (values in the chart are not yearly).

When broken down to geographical regions, Table (A.1) shows that Google is dominant in Europe and Latin America, but not in Asia-Pacific region. This is mainly because Baidu.com is the number one search engine in China and Yahoo in Japan. In Europe, only few countries deviate from the overall trend – Russia with Yandex and the Czech Republic with Seznam.cz.

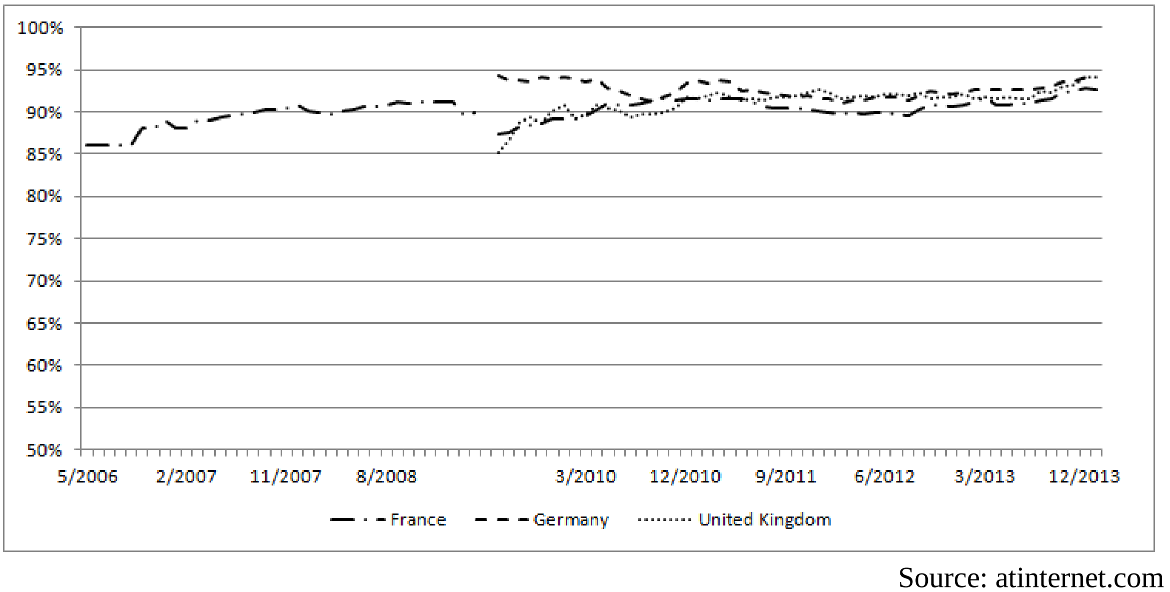
**Table A.1: Share of search engines – by geographical region**

| | Asia-Pacific | | Latin America | Europe | | |
|---|---|---|---|---|---|---|
| | 4/2008 | 7/2008 | 3/2011 | 3/2008 | 1/2011 | 2/2013 |
| Google Sites | 39% | 34% | 91% | 79% | 90% | . |
| Yahoo! Sites | 24% | 20% | 1% | 2% | . | . |
| Microsoft Sites | 2% | 2% | 3% | 2% | . | . |
| Baidu.com Inc. | 17% | 27% | . | . | . | . |
| Yandex | . | . | . | 2% | . | 10% |

Source: comScore.com

Figure (A.11) shows the popularity of Google in France, Germany, and the United Kingdom. These values express a little different phenomenon than those presented before – comScore.com provided number about shares of searches, the values below from atinternet.com rather measure the traffic on individual tracked websites that is directed to them from a particular search engine.

**Figure A.11: Share of Google in France, Germany and the United Kingdom**

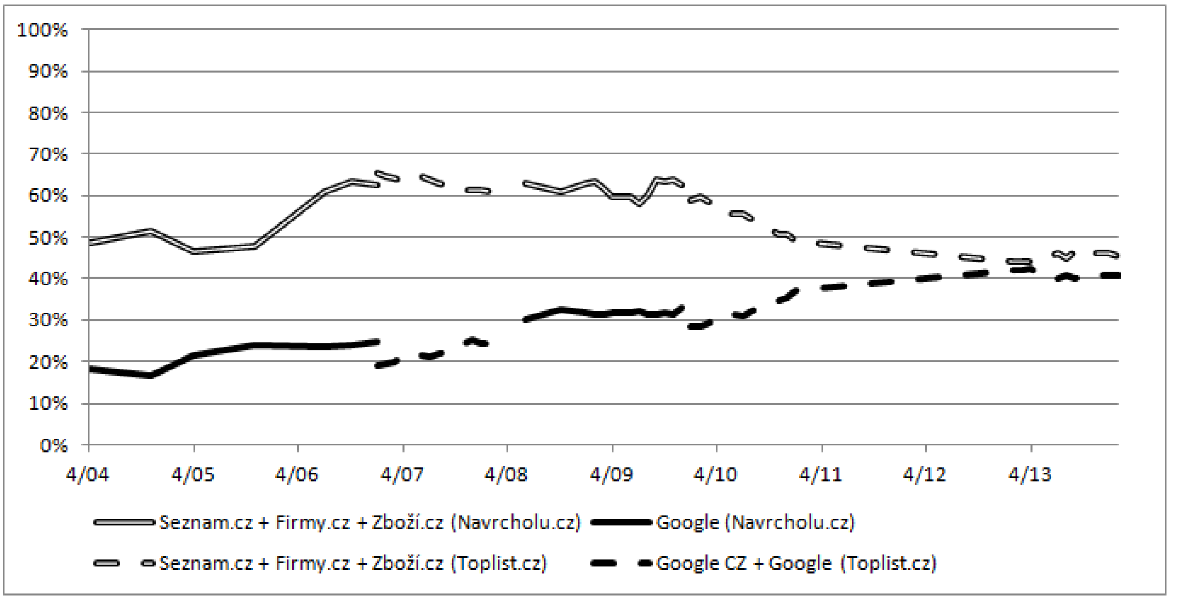



Source: atinternet.com

The position of Google was clearly superior to other search engines. On the other hand, the position of Google in the Czech Republic was inferior to Seznam.cz for most of the analyzed period, as is depicted in Figure (A.12).

The values provided are a combination from two sources, and again it is a measure of traffic on tracked websites that was directed to them from a particular search engine.

**Figure A.12: Share of Google in the Czech Republic (2004–2013)**




Source: navrcholu.cz, toplist.cz

The share of Google was below 20% in 2004 and reached 40% only by 2012, slowly closing the gap to the first place. In 2004, Google did not even have the local domain (Google.cz, acquired in September 2006), people wishing to use its services had to use either Google.com or other national mutations (such as Google.de). Google further targeted potential customer in the Czech Republic by full localization that started in 2008 and an excessive TV campaign in 2010, which seemed to affect the share of Google around that year.

The values presented may underestimate the real share of Google because of the methodology of measurement. Firstly, only a sample of Czech websites is tracked by both source (navrcholu.cz and toplist.cz), and since Seznam.cz directs its users only to Czech websites (in contrast to Google), the actual share of searches is probably different. Also, internet users differ in how many pages they open after one search (this is also affected by the quality of suggestions provided); so in overall, this ”click trough” method of measurement does not perfectly capture the share of searches conducted.

# Appendix B: Additional tables and figures

**Table B.1: Out-of-sample nowcasting of unemployment – combination with the current value of Share of unemployed**

| Out-of-sample period | | 2009 - 2013 | 2010 - 2013 | 2011 - 2013 |
|---|---|---|---|---|
| Autoregressive model AR(1) | MSE | 0.0005237 | 0.0002571 | 0.0002887 |
| AR(1) with the current Share of Unemployed (ARX) | | 95.8% | 111.5% | 105.4% |
| Additional explanatory variable | Lag | % MSE of ARX | % MSE of ARX | % MSE of ARX |
| Google query: "job" | 0 | **97.8%**<br>1.97<br>(0.027)<br>** | **95.0%**<br>2.01<br>(0.025)<br>** | **93.1%**<br>2.19<br>(0.018)<br>** |
| Google query: "employment" | 0 | **99.2%**<br>1.16<br>(0.125) | **97.3%**<br>1.83<br>(0.037)<br>** | **96.3%**<br>2.14<br>(0.020)<br>** |
| Google query: "job offers" | 0 | **97.8%**<br>1.12<br>(0.133) | **96.7%**<br>0.90<br>(0.187) | **95.6%**<br>0.92<br>(0.182) |
| Google query: "Labour office" | 2 | **97.2%**<br>2.16<br>(0.017)<br>** | **93.7%**<br>2.23<br>(0.015)<br>** | **93.1%**<br>2.11<br>(0.021)<br>** |
| Google query: "CV" | 4 | **95.8%**<br>2.08<br>(0.021)<br>** | **94.1%**<br>1.72<br>(0.046)<br>** | **92.9%**<br>1.68<br>(0.051)<br>* |

Source: author's calculations

Explanation: The percentage value shows the relative MSE of the augmented ARX model to the benchmark ARX model, value lower than 100% implies improved forecasts when additional variable was introduced. Next rows show the Clark-West test statistics, p-value and significance of this test. The null hypothesis is equal predictive accuracy of both models; we cannot reject it for high p-values, meaning that potential improvement is not statistically significant. Alternative hypothesis is superior forecasting quality of the larger model. Significance is denoted by asterisks (* for 10%, ** for 5% and *** for 1% significance level).

**Table B.2: Results of logit models for economic downturns (lags 0–3 for quarterly data)**

| Lag | | GCSI | BCI | CCI | CLIs | PX | ER |
|---|---|---|---|---|---|---|---|
| | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -13.5 *** | -16.1 ** | -9.7 *** | -13.5 *** | -12.3 *** | -18.0 |
| 0 | McFadden $R^2$ | **0.282** | **0.143** | **0.484** | **0.280** | **0.346** | **0.043** |
| | % of correct forecasts | **82%** | **64%** | **89%** | **79%** | **82%** | **61%** |
| | % of correct downturn fc | **73%** | **55%** | **91%** | **64%** | **82%** | **18%** |
| | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -13.7 *** | -17.9 | -13.5 *** | -17.1 | -16.3 ** | -15.4 ** |
| 1 | McFadden $R^2$ | **0.251** | **0.017** | **0.262** | **0.060** | **0.109** | **0.155** |
| | % of correct forecasts | **82%** | **56%** | **78%** | **59%** | **70%** | **74%** |
| | % of correct downturn fc | **73%** | **9%** | **73%** | **18%** | **64%** | **64%** |
| | Sign of Beta | - | + | - | + | - | - |
| | Log-likelihood | -15.3 ** | -17.3 | -16.3 * | -17.7 | -17.5 | -10.1 *** |
| 2 | McFadden $R^2$ | **0.138** | **0.022** | **0.077** | **0.003** | **0.015** | **0.430** |
| | % of correct forecasts | **69%** | **50%** | **62%** | **58%** | **54%** | **77%** |
| | % of correct downturn fc | **64%** | **18%** | **45%** | **0%** | **0%** | **73%** |
| | Sign of Beta | - | + | - | + | + | - |
| | Log-likelihood | -16.0 | -15.0 ** | -17.0 | -15.3 * | -16.9 | -12.4 *** |
| 3 | McFadden $R^2$ | **0.070** | **0.123** | **0.006** | **0.105** | **0.012** | **0.277** |
| | % of correct forecasts | **64%** | **68%** | **56%** | **64%** | **56%** | **80%** |
| | % of correct downturn fc | **55%** | **55%** | **18%** | **55%** | **27%** | **82%** |

Source: author's calculations

Explanation: Results are presented for the set of 6 explanatory variables: GCSI (Google Consumer Sentiment Index), BCI (Business Confidence Indicator), CCI (Consumer Confidence Indicator), CLIs (Composite Leading Indicators), PX (Prague Stock Exchange Index), ER (CZK/EUR exchange rate).

Each cell contains results of one logit model – each column shows results for a particular explanatory variable, each row shows results for a particular lag of given variables when explaining current state of the economy (downturn).

Each cell contains the following: (1) sign of beta estimate: positive beta means that higher value of explanatory variable is connected with a higher probability of downturn (and vice versa); (2) log-likelihood and significance of the model based on likelihood ratio; (3) McFadden $R^2$ measuring fit of the model; (4) percentage of correct forecasts; (5) percentage of correct downturn forecasts (from all downturns).

**Table B.3: Results of logit models for below-average growth of household consumption (lags 0–3 for quarterly data)**

| Lag | | GCSI | BCI | CCI | CLIs | PX | U |
|---|---|---|---|---|---|---|---|
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -10.2 *** | -16.0 ** | -12.7 *** | -16.8 ** | -14.9 *** | -16.7 ** |
| 0 | McFadden $R^2$ | **0.469** | **0.161** | **0.336** | **0.122** | **0.222** | **0.125** |
| | % of correct forecasts | **89%** | **54%** | **79%** | **61%** | **68%** | **57%** |
| | % of correct downturn fc | **83%** | **42%** | **75%** | **42%** | **67%** | **67%** |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -11.4 *** | -16.6 ** | -14.2 *** | -16.6 ** | -14.0 *** | -17.8 |
| 1 | McFadden $R^2$ | **0.388** | **0.104** | **0.237** | **0.104** | **0.245** | **0.042** |
| | % of correct forecasts | **78%** | **52%** | **78%** | **52%** | **70%** | **48%** |
| | % of correct downturn fc | **75%** | **42%** | **75%** | **33%** | **67%** | **50%** |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -14.1 *** | -17.2 | -16.5 * | -17.4 | -15.6 ** | -17.8 |
| 2 | McFadden $R^2$ | **0.214** | **0.041** | **0.082** | **0.029** | **0.129** | **0.008** |
| | % of correct forecasts | **65%** | **46%** | **62%** | **58%** | **62%** | **42%** |
| | % of correct downturn fc | **58%** | **33%** | **58%** | **25%** | **58%** | **17%** |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -15.7 * | -17.3 | -17.2 | -17.3 | -16.9 | -17.3 |
| 3 | McFadden $R^2$ | **0.092** | **0.003** | **0.008** | **0.000** | **0.023** | **0.000** |
| | % of correct forecasts | **56%** | **36%** | **48%** | **52%** | **52%** | **52%** |
| | % of correct downturn fc | **58%** | **17%** | **42%** | **0%** | **50%** | **0%** |

Source: author's calculations

Explanation: Results are presented for the set of 6 explanatory variables: GCSI (Google Consumer Sentiment Index), BCI (Business Confidence Indicator), CCI (Consumer Confidence Indicator), CLIs (Composite Leading Indicators), PX (Prague Stock Exchange Index), U (Unemployment Rate).

Each cell contains results of one logit model – each column shows results for a particular explanatory variable, each row shows results for a particular lag of given variables when explaining current below-average growth of household consumption.

Each cell contains the following: (1) sign of beta estimate: positive beta means that higher value of explanatory variable is connected with a higher probability of downturn (and vice versa); (2) log-likelihood and significance of the model based on likelihood ratio; (3) McFadden $R^2$ measuring fit of the model; (4) percentage of correct forecasts; (5) percentage of correct downturn forecasts (from all downturns).

**Table B.4: Results of logit models for below-average growth of household consumption before the elimination of one-off changes (lags 0–5 for monthly data)**

| Lag | | GCSI | BCI | CCI | CLIs | PX | U |
|---|---|---|---|---|---|---|---|
| 0 | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -45.7 *** | -53.4 *** | -49.9 *** | -54.7 *** | -53.6 *** | -53.8 *** |
| | McFadden $R^2$ | **0.212** | **0.079** | **0.140** | **0.058** | **0.077** | **0.073** |
| | % of correct forecasts | **74%** | **52%** | **68%** | **57%** | **66%** | **63%** |
| | % of correct downturn fc | **74%** | **44%** | **67%** | **41%** | **64%** | **74%** |
| 1 | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -45.9 *** | -52.9 *** | -49.3 *** | -53.6 *** | -52.6 *** | -54.3 ** |
| | McFadden $R^2$ | **0.199** | **0.078** | **0.140** | **0.065** | **0.084** | **0.054** |
| | % of correct forecasts | **74%** | **60%** | **70%** | **53%** | **63%** | **61%** |
| | % of correct downturn fc | **74%** | **56%** | **69%** | **41%** | **64%** | **72%** |
| 2 | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -48.2 *** | -52.6 *** | -49.9 *** | -53.0 *** | -51.0 *** | -54.6 ** |
| | McFadden $R^2$ | **0.150** | **0.073** | **0.121** | **0.066** | **0.101** | **0.038** |
| | % of correct forecasts | **70%** | **63%** | **67%** | **52%** | **62%** | **59%** |
| | % of correct downturn fc | **72%** | **62%** | **67%** | **41%** | **64%** | **69%** |
| 3 | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -49.0 *** | -52.5 *** | -50.0 *** | -52.7 *** | -49.0 *** | -54.8 |
| | McFadden $R^2$ | **0.126** | **0.065** | **0.108** | **0.060** | **0.126** | **0.023** |
| | % of correct forecasts | **70%** | **62%** | **68%** | **52%** | **62%** | **54%** |
| | % of correct downturn fc | **72%** | **59%** | **69%** | **41%** | **67%** | **64%** |
| 4 | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -49.5 *** | -52.4 ** | -51.3 *** | -52.8 ** | -48.8 *** | -54.8 |
| | McFadden $R^2$ | **0.106** | **0.054** | **0.075** | **0.047** | **0.120** | **0.012** |
| | % of correct forecasts | **66%** | **61%** | **65%** | **54%** | **58%** | **53%** |
| | % of correct downturn fc | **67%** | **59%** | **67%** | **46%** | **64%** | **64%** |
| 5 | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -50.7 *** | -52.6 ** | -52.0 ** | -53.0 * | -49.3 *** | -54.5 |
| | McFadden $R^2$ | **0.075** | **0.040** | **0.051** | **0.031** | **0.100** | **0.005** |
| | % of correct forecasts | **58%** | **62%** | **61%** | **57%** | **58%** | **54%** |
| | % of correct downturn fc | **59%** | **62%** | **62%** | **51%** | **67%** | **69%** |

Source: author's calculations

Explanation: Results are presented for the set of 6 explanatory variables: GCSI (Google Consumer Sentiment Index), BCI (Business Confidence Indicator), CCI (Consumer Confidence Indicator), CLIs (Composite Leading Indicators), PX (Prague Stock Exchange Index), U (Unemployment Rate).

Each cell contains results of one logit model – each column shows results for a particular explanatory variable, each row shows results for a particular lag of given variables when explaining current below-average growth of household consumption before the elimination of one-off changes.

Each cell contains the following: (1) sign of beta estimate: positive beta means that higher value of explanatory variable is connected with a higher probability of downturn (and vice versa); (2) log-likelihood and significance of the model based on likelihood ratio; (3) McFadden $R^2$ measuring fit of the model; (4) percentage of correct forecasts; (5) percentage of correct downturn forecasts (from all downturns).

**Table B.5: Results of logit models for below-average growth of household consumption before the elimination of one-off changes (lags 6–11 for monthly data)**

| Lag | | GCSI | BCI | CCI | CLIs | PX | U |
|---|---|---|---|---|---|---|---|
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -51.4 ** | -52.4 * | -52.1 ** | -53.2 | -49.9 *** | -54.0 |
| 6 | McFadden $R^2$ | **0.049** | **0.030** | **0.036** | **0.017** | **0.077** | **0.002** |
| | % of correct forecasts | **51%** | **56%** | **56%** | **58%** | **60%** | **55%** |
| | % of correct downturn fc | **51%** | **59%** | **56%** | **54%** | **72%** | **69%** |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -51.6 * | -52.5 | -52.5 | -53.0 | -50.5 ** | -53.3 |
| 7 | McFadden $R^2$ | **0.033** | **0.015** | **0.017** | **0.006** | **0.054** | **0.001** |
| | % of correct forecasts | **56%** | **55%** | **53%** | **58%** | **61%** | **52%** |
| | % of correct downturn fc | **62%** | **59%** | **54%** | **62%** | **74%** | **69%** |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -51.5 | -52.4 | -52.3 | -52.6 | -51.3 | -52.6 |
| 8 | McFadden $R^2$ | **0.022** | **0.005** | **0.006** | **0.001** | **0.025** | **0.000** |
| | % of correct forecasts | **59%** | **59%** | **55%** | **59%** | **63%** | **45%** |
| | % of correct downturn fc | **69%** | **77%** | **59%** | **85%** | **82%** | **82%** |
| | Sign of Beta | - | - | - | + | - | + |
| | Log-likelihood | -51.3 | -51.9 | -51.9 | -51.9 | -51.4 | -51.9 |
| 9 | McFadden $R^2$ | **0.012** | **0.001** | **0.001** | **0.000** | **0.011** | **0.000** |
| | % of correct forecasts | **56%** | **64%** | **59%** | **52%** | **57%** | **52%** |
| | % of correct downturn fc | **72%** | **92%** | **92%** | **100%** | **79%** | **100%** |
| | Sign of Beta | - | - | - | + | - | + |
| | Log-likelihood | -50.4 | -51.2 | -51.2 | -51.3 | -50.8 | -51.3 |
| 10 | McFadden $R^2$ | **0.017** | **0.001** | **0.001** | **0.000** | **0.010** | **0.000** |
| | % of correct forecasts | **61%** | **58%** | **55%** | **54%** | **57%** | **51%** |
| | % of correct downturn fc | **71%** | **82%** | **71%** | **100%** | **79%** | **100%** |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -48.8 * | -50.5 | -50.5 | -50.6 | -50.2 | -50.6 |
| 11 | McFadden $R^2$ | **0.036** | **0.001** | **0.002** | **0.000** | **0.007** | **0.001** |
| | % of correct forecasts | **62%** | **58%** | **60%** | **51%** | **58%** | **59%** |
| | % of correct downturn fc | **68%** | **76%** | **62%** | **100%** | **78%** | **76%** |

Source: author's calculations

Explanation: Results are presented for the set of 6 explanatory variables: GCSI (Google Consumer Sentiment Index), BCI (Business Confidence Indicator), CCI (Consumer Confidence Indicator), CLIs (Composite Leading Indicators), PX (Prague Stock Exchange Index), U (Unemployment Rate).

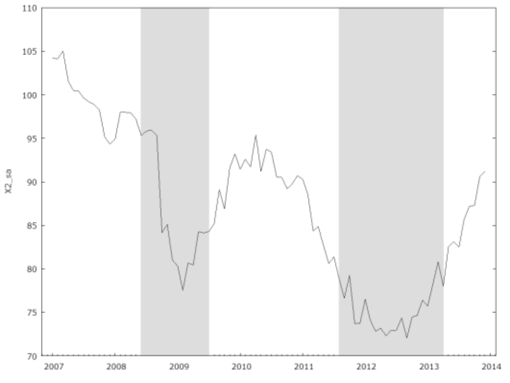
Each cell contains results of one logit model – each column shows results for a particular explanatory variable, each row shows results for a particular lag of given variables when explaining current below-average growth of household consumption before the elimination of one-off changes.

Each cell contains the following: (1) sign of beta estimate: positive beta means that higher value of explanatory variable is connected with a higher probability of downturn (and vice versa); (2) log-likelihood and significance of the model based on likelihood ratio; (3) McFadden $R^2$ measuring fit of the model; (4) percentage of correct forecasts; (5) percentage of correct downturn forecasts (from all downturns).

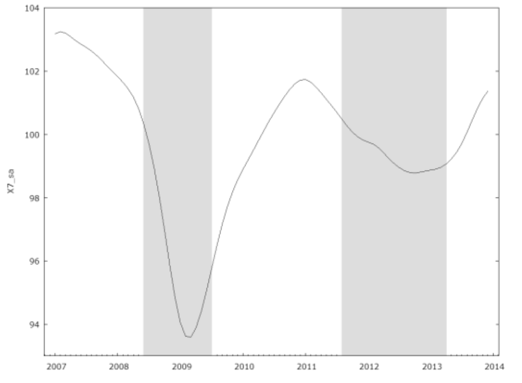
**Figure B.1: Individual explanatory variables and periods of economic downturn**

Google Consumer Sentiment Index (GCSI)

Business Confidence Indicator (BCI)

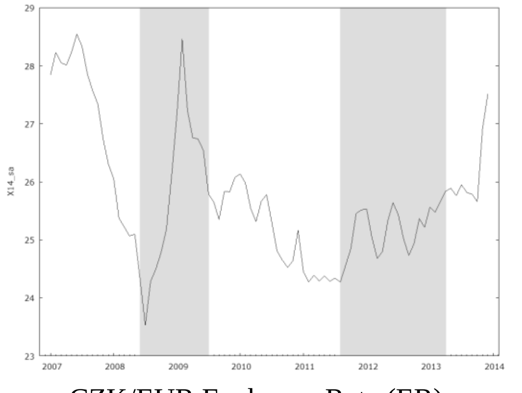


Consumer Confidence Indicator (CCI)



Composite Leading Indicators (CLIs)

Prague Stock Exchange Index (PX)



CZK/EUR Exchange Rate (ER)

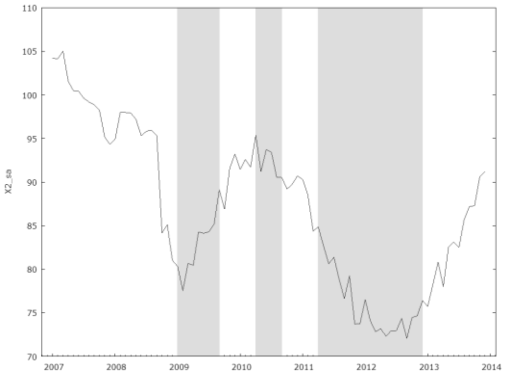
Source: CZSO, OECD, CNB, Prague Stock Exchange; author's calculations

Explanation: The figures depict the development of individual explanatory variables over the period 2007–2013 compared to periods of economic downturns (shaded area) as defined by OECD (2014).

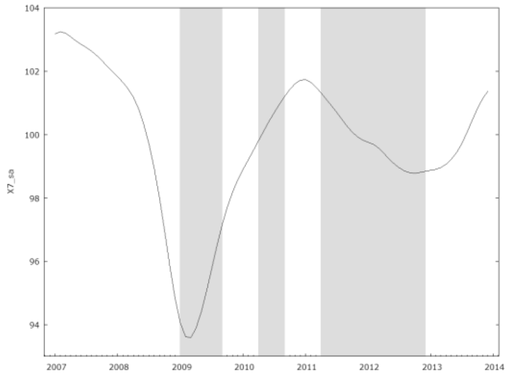
**Figure B.2: Individual explanatory variables and periods of below average growth of household consumption**

Google Consumer Sentiment Index (GCSI)

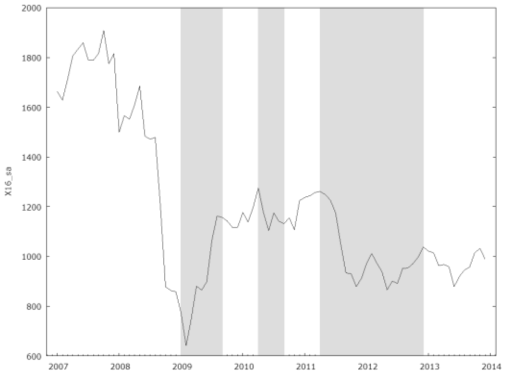
Business Confidence Indicator (BCI)



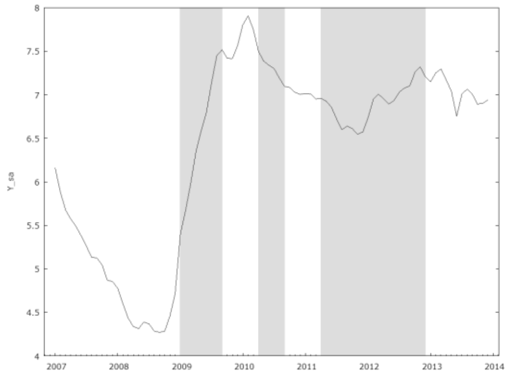
Consumer Confidence Indicator (CCI)



Composite Leading Indicators (CLIs)



Prague Stock Exchange Index (PX)

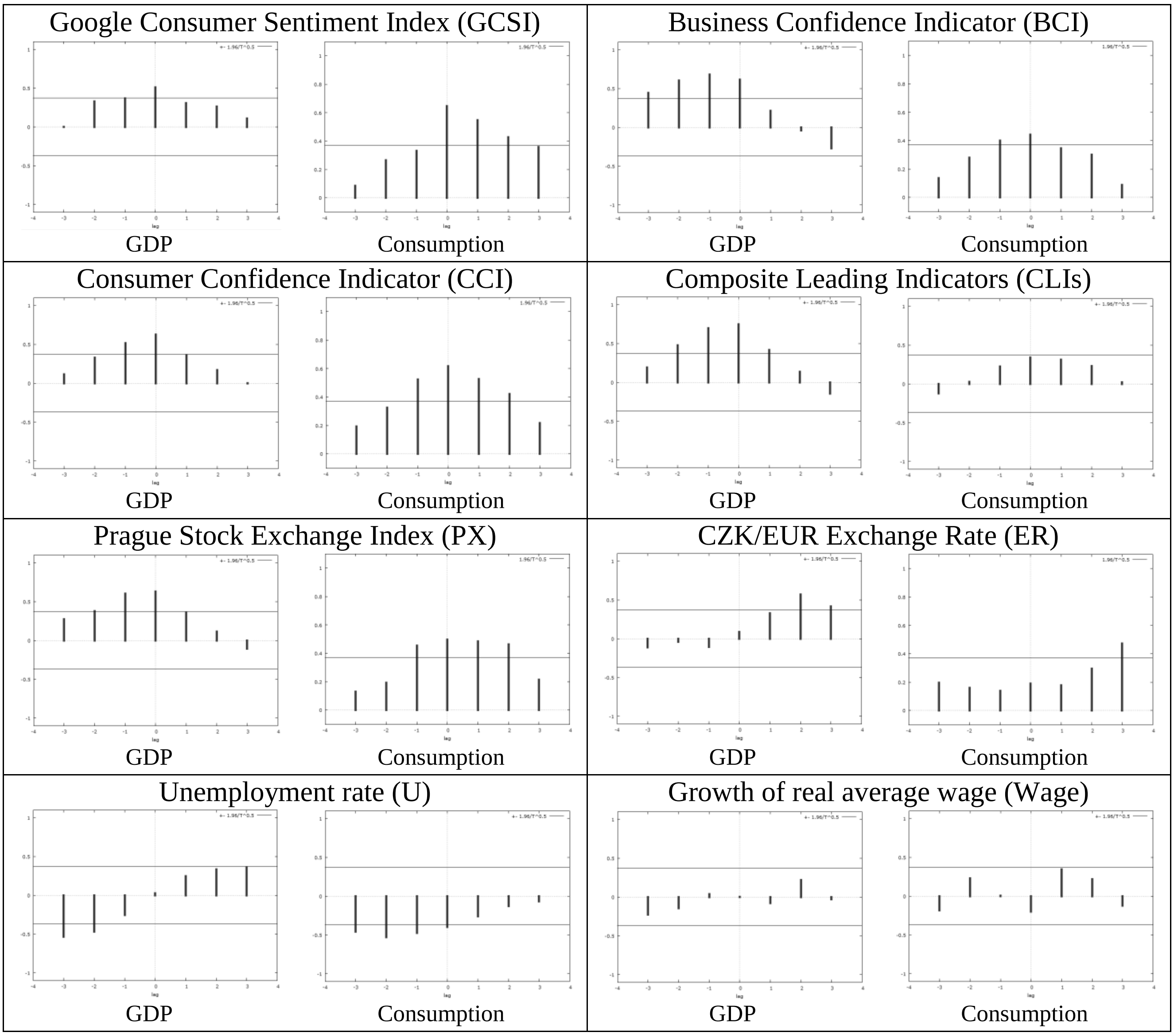


Unemployment rate (U)

Source: CZSO, OECD, CNB, Prague Stock Exchange; author's calculations

Explanation: The figures depict the development of individual explanatory variables over the period 2007–2013 compared to periods of below average growth of household consumption (shaded area).

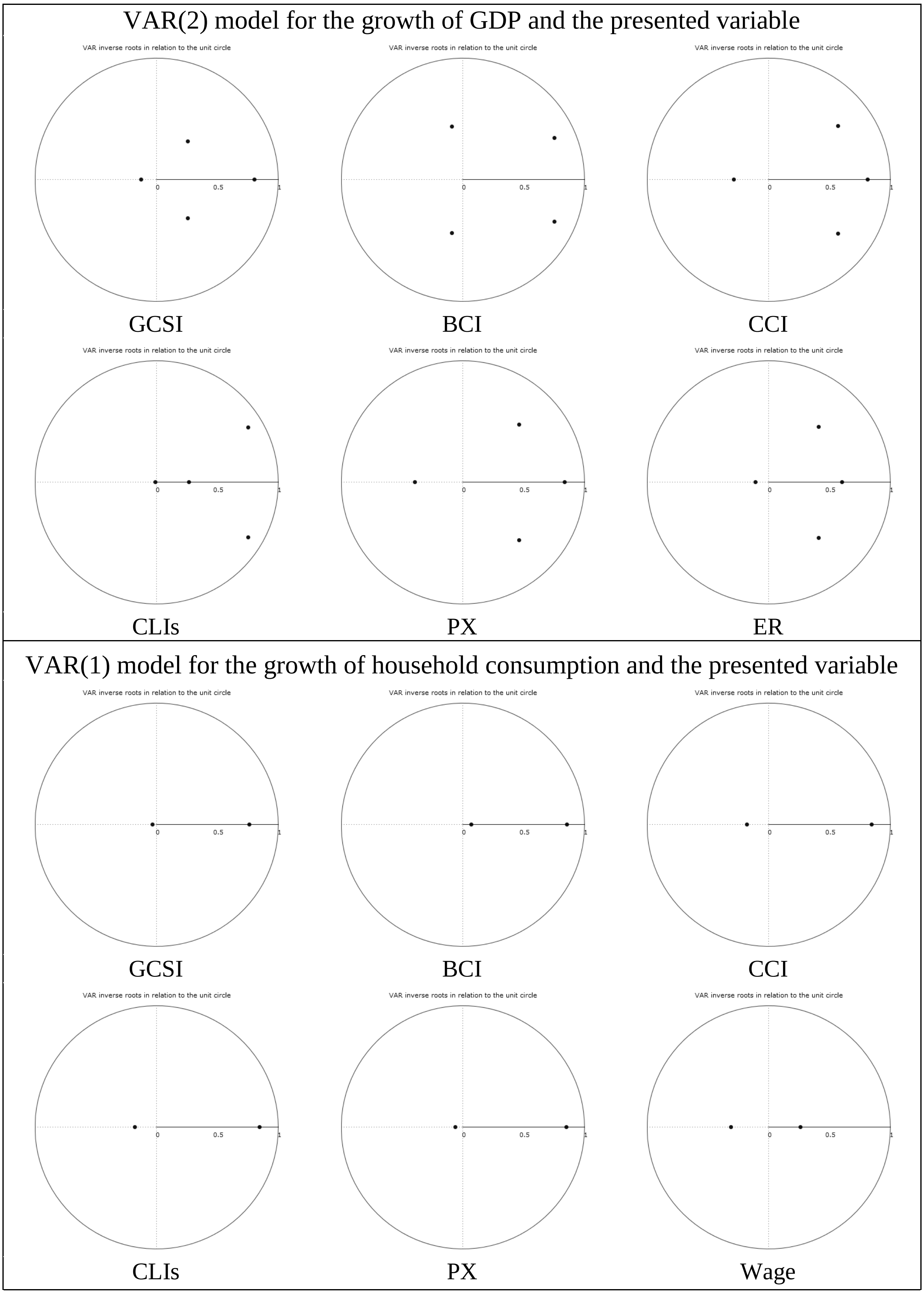
**Figure B.3: Cross-correlograms between individual variables and quarterly growth of GDP or household consumption**




Source: author's calculations in Gretl

Explanation: The charts show cross-correlograms between individual explanatory variables and either GDP or household consumption growth. In each chart, the LHS shows values of correlation between lagged values of GDP/consumption and the current values of explanatory variable; the RHS shows values of correlation between lagged values of explanatory variables and the current value of GDP/consumption. Higher values on the RHS would indicate that the explanatory variable is the leading one.

**Figure B.4: Bivariate VAR unit root test – stability of models**

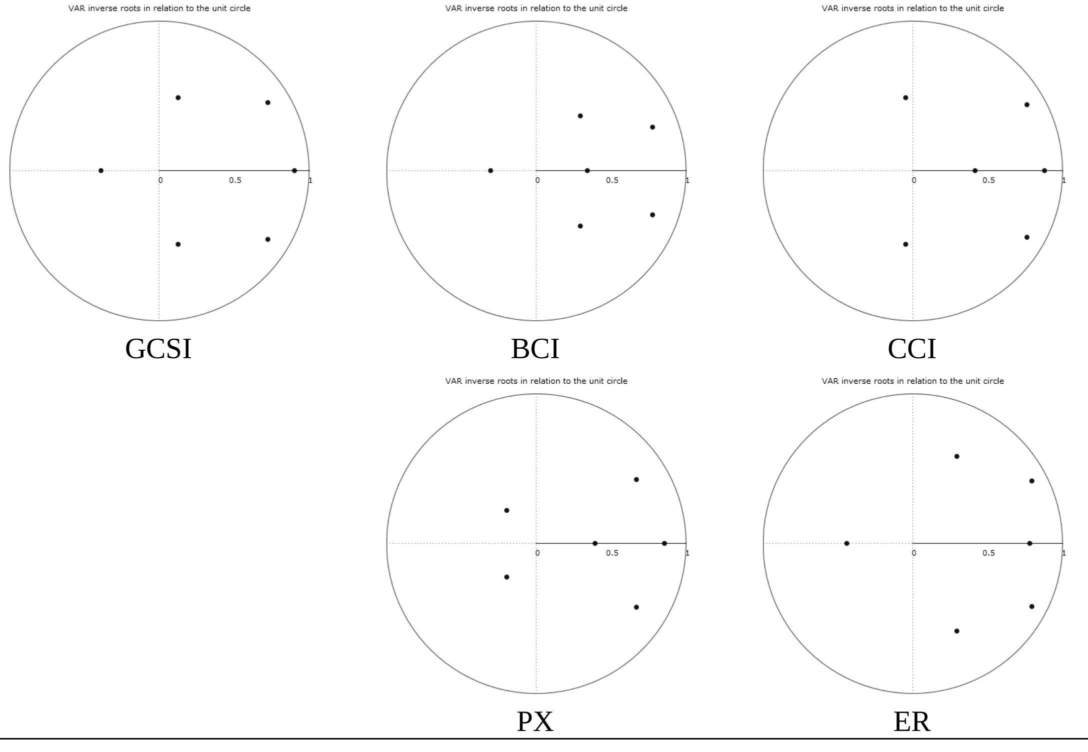



Source: author's calculations in Gretl

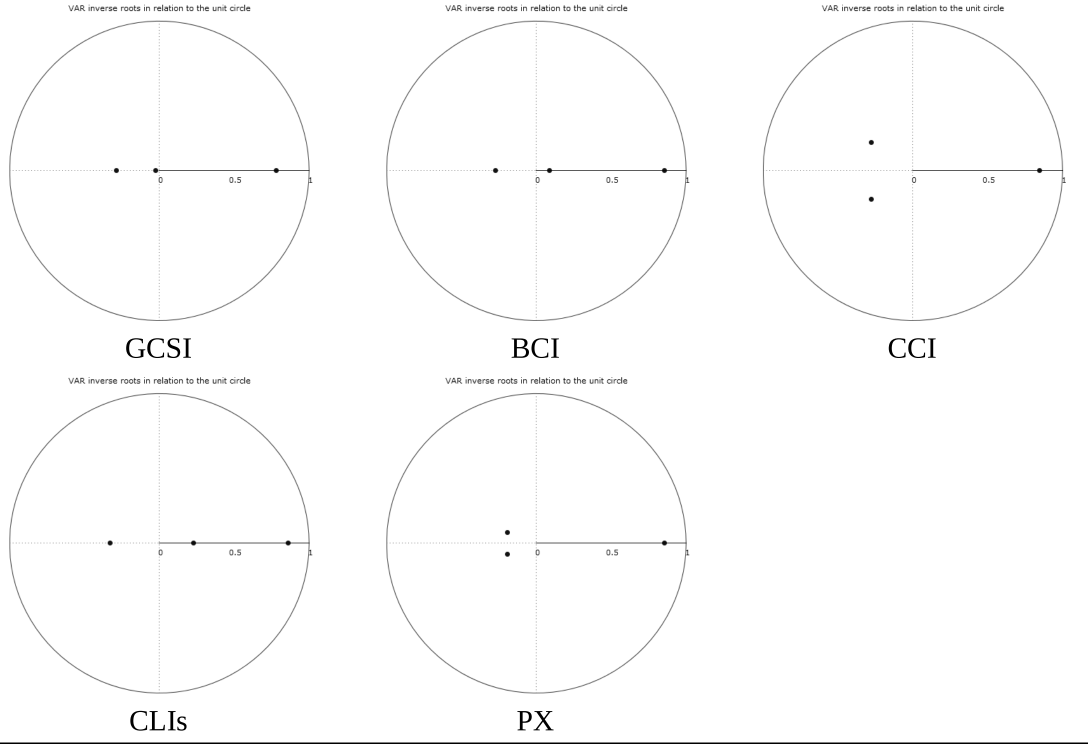
Explanation: Figures – unit root test for bivariate models estimated over the whole sample 2007–2013 – are provided for the following variables: Google Consumer Sentiment Index (GCSI), Business Confidence Indicator (BCI), Consumer Confidence Indicator (CCI), Composite Leading Indicators (CLIs), Prague Stock Exchange Index (PX), CZK/EUR Exchange Rate (ER), growth of real average wage (Wage).

**Figure B.5: Trivariate VAR unit root test – stability of models**

VAR(2) model for the growth of GDP, CLIs, and the presented variable




VAR(1) model for the growth of h. consumption, Wage, and the presented variable




Source: author's calculations in Gretl

Explanation: Figures – unit root test for trivariate models estimated over the whole sample 2007–2013 – are provided for the following variables: Google Consumer Sentiment Index (GCSI), Business Confidence Indicator (BCI), Consumer Confidence Indicator (CCI), Composite Leading Indicators (CLIs), Prague Stock Exchange Index (PX), CZK/EUR Exchange Rate (ER), growth of real average wage (Wage).

# Appendix C: Results revisited

**Table C.1: Out-of-sample nowcasting of unemployment (various periods)**

| Out-of-sample period | | 2009 - 2014 | 2010 - 2014 | 2011 - 2014 |
|---|---|---|---|---|
| Autoregressive model AR2 (benchmark) | MSE | 0.0005505 | 0.0003879 | 0.0003893 |
| Additional explanatory variable | Lag | % MSE of AR2 | % MSE of AR2 | % MSE of AR2 |
| Consumer Confidence Indicator | 7 | **100.0%**<br>1.56<br>(0.061)<br>* | **101.7%**<br>0.84<br>(0.201) | **105.0%**<br>-0.03<br>---- |
| Composite Confidence Indicator | 7 | **101.5%**<br>1.45<br>(0.076)<br>* | **101.8%**<br>0.96<br>(0.171) | **100.8%**<br>0.80<br>(0.213) |
| Index of Industrial Production | 8 | **93.5%**<br>2.12<br>(0.019)<br>** | **97.8%**<br>1.22<br>(0.113) | **98.4%**<br>0.88<br>(0.192) |
| Composite Leading Indicators | 8 | **106.0%**<br>0.53<br>(0.297) | **99.0%**<br>0.66<br>(0.257) | **97.4%**<br>0.87<br>(0.194) |
| Share of Unemployed | 2 | **112.6%**<br>0.82<br>(0.206) | **101.1%**<br>1.98<br>(0.026)<br>** | **102.8%**<br>1.81<br>(0.039)<br>** |
| Google query: "job" | 0 | **99.8%**<br>0.77<br>(0.221) | **99.9%**<br>0.74<br>(0.230) | **97.4%**<br>1.06<br>(0.148) |
| Google query: "employment" | 0 | **99.8%**<br>0.30<br>(0.384) | **100.4%**<br>-0.31<br>---- | **101.0%**<br>-1.19<br>---- |
| Google query: "job offers" | 0 | **97.7%**<br>1.94<br>(0.028)<br>** | **96.0%**<br>2.24<br>(0.015)<br>** | **95.0%**<br>2.30<br>(0.013)<br>** |
| Google query: "Labour office" | 2 | **101.0%**<br>0.53<br>(0.298) | **101.6%**<br>0.57<br>(0.287) | **103.3%**<br>0.31<br>(0.380) |
| Google query: "CV" | 4 | **98.8%**<br>0.91<br>(0.183) | **98.9%**<br>0.77<br>(0.221) | **98.4%**<br>0.89<br>(0.188) |

Source: author's calculations

Explanation: The percentage value shows the relative MSE of the augmented ARX model to the benchmark AR(2) model, value lower than 100% implies improved forecasts when additional variable was introduced. Next rows show the Clark-West test statistics, p-value and significance of this test. The null hypothesis is equal predictive accuracy of both models; we cannot reject it for high p-values, meaning that potential improvement is not statistically significant. Alternative hypothesis is superior forecasting quality of the larger model. Significance is denoted by asterisks (* for 10%, ** for 5%, and *** for 1% significance level).

**Table C.2: Out-of-sample nowcasting of unemployment – combination with the Index of Industrial Production**

| Out-of-sample period | | 2009 - 2014 | 2010 - 2014 | 2011 - 2014 |
|---|---|---|---|---|
| Autoregressive model AR(2) | MSE | 0.0005505 | 0.0003879 | 0.0003893 |
| AR(2) with 8th lag of Index of Industrial Production (ARX) | | 93.5% | 97.8% | 98.4% |
| Additional explanatory variable | Lag | % MSE of ARX | % MSE of ARX | % MSE of ARX |
| Consumer Confidence Indicator | 7 | **99.6%**<br>1.80<br>(0.038)<br>** | **101.5%**<br>0.90<br>(0.186) | **105.0%**<br>0.00<br>(0.500) |
| Composite Confidence Indicator | 7 | **104.3%**<br>1.15<br>(0.128) | **101.0%**<br>0.92<br>(0.181) | **99.7%**<br>0.84<br>(0.202) |
| Composite Leading Indicators | 8 | **107.6%**<br>0.27<br>(0.395) | **99.6%**<br>0.50<br>(0.309) | **98.0%**<br>0.81<br>(0.210) |
| Share of Unemployed | 2 | **116.4%**<br>0.70<br>(0.242) | **100.8%**<br>1.97<br>(0.027)<br>** | **103.1%**<br>1.73<br>(0.046)<br>** |
| Google query: "job" | 0 | **99.2%**<br>1.01<br>(0.159) | **98.6%**<br>0.99<br>(0.163) | **96.8%**<br>1.21<br>(0.117) |
| Google query: "employment" | 0 | **101.1%**<br>-1.31<br>---- | **100.6%**<br>-0.64<br>---- | **101.1%**<br>-1.13<br>---- |
| Google query: "job offers" | 0 | **96.1%**<br>2.43<br>(0.009)<br>*** | **95.8%**<br>2.17<br>(0.017)<br>** | **94.6%**<br>2.23<br>(0.015)<br>** |
| Google query: "Labour office" | 2 | **103.0%**<br>0.03<br>(0.486) | **104.2%**<br>0.17<br>(0.433) | **106.3%**<br>-0.01<br>---- |
| Google query: "CV" | 4 | **99.3%**<br>0.83<br>(0.206) | **99.4%**<br>0.64<br>(0.261) | **99.0%**<br>0.74<br>(0.231) |

Source: author's calculations

Explanation: The percentage value shows the relative MSE of the augmented ARX model to the benchmark ARX model, value lower than 100% implies improved forecasts when additional variable was introduced. Next rows show the Clark-West test statistics, p-value and significance of this test. The null hypothesis is equal predictive accuracy of both models; we cannot reject it for high p-values, meaning that potential improvement is not statistically significant. Alternative hypothesis is superior forecasting quality of the larger model. Significance is denoted by asterisks (* for 10%, ** for 5%, and *** for 1% significance level).

**Table C.3: Out-of-sample nowcasting of unemployment – combination with the Composite Leading Indicators**

| Out-of-sample period | | 2009 - 2014 | 2010 - 2014 | 2011 - 2014 |
|---|---|---|---|---|
| Autoregressive model AR(2) | MSE | 0.0005505 | 0.0003879 | 0.0003893 |
| AR(2) with 8th lag of Composite Leading Indicators (ARX) | | 106.0% | 99.0% | 97.4% |
| Additional explanatory variable | Lag | % MSE of ARX | % MSE of ARX | % MSE of ARX |
| Consumer Confidence Indicator | 7 | **101.0%**<br>0.63<br>(0.264) | **100.5%**<br>0.94<br>(0.176) | **104.0%**<br>-0.22<br>---- |
| Composite Confidence Indicator | 7 | **102.5%**<br>0.91<br>(0.183) | **103.0%**<br>0.58<br>(0.281) | **101.5%**<br>0.47<br>(0.319) |
| Index of Industrial Production | 8 | **95.0%**<br>2.71<br>(0.004)<br>*** | **98.4%**<br>0.91<br>(0.185) | **99.0%**<br>0.67<br>(0.253) |
| Share of Unemployed | 2 | **114.8%**<br>0.08<br>(0.469) | **99.6%**<br>1.79<br>(0.039)<br>** | **102.2%**<br>1.49<br>(0.072)<br>* |
| Google query: "job" | 0 | **98.8%**<br>1.08<br>(0.142) | **100.2%**<br>0.73<br>(0.235) | **97.3%**<br>1.14<br>(0.131) |
| Google query: "employment" | 0 | **100.9%**<br>-1.07<br>---- | **100.3%**<br>-0.27<br>---- | **100.9%**<br>-1.16<br>---- |
| Google query: "job offers" | 0 | **96.9%**<br>2.37<br>(0.010)<br>** | **95.6%**<br>2.42<br>(0.009)<br>*** | **94.5%**<br>2.45<br>(0.009)<br>*** |
| Google query: "Labour office" | 2 | **101.4%**<br>0.38<br>(0.354) | **100.7%**<br>0.81<br>(0.212) | **102.3%**<br>0.54<br>(0.297) |
| Google query: "CV" | 4 | **98.3%**<br>1.53<br>(0.066)<br>* | **97.8%**<br>1.25<br>(0.107) | **97.0%**<br>1.33<br>(0.095)<br>* |

Source: author's calculations

Explanation: The percentage value shows the relative MSE of the augmented ARX model to the benchmark ARX model, value lower than 100% implies improved forecasts when additional variable was introduced. Next rows show the Clark-West test statistics, p-value and significance of this test. The null hypothesis is equal predictive accuracy of both models; we cannot reject it for high p-values, meaning that potential improvement is not statistically significant. Alternative hypothesis is superior forecasting quality of the larger model. Significance is denoted by asterisks (* for 10%, ** for 5%, and *** for 1% significance level).

**Table C.4: Results of logit models for economic downturns (lags 0–5 for monthly data)**

| Lag | | GCSI | BCI | CCI | CLIs | PX | ER |
|---|---|---|---|---|---|---|---|
| | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -46.9 *** | -52.6 *** | -34.1 *** | -39.7 *** | -49.0 *** | -57.7 *** |
| 0 | McFadden $R^2$ | **0.255** | **0.164** | **0.458** | **0.370** | **0.222** | **0.083** |
| | % of correct forecasts | **73%** | **73%** | **82%** | **80%** | **78%** | **64%** |
| | % of correct downturn fc | **57%** | **54%** | **77%** | **69%** | **69%** | **31%** |
| | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -47.5 *** | -56.1 *** | -38.0 *** | -45.6 *** | -52.5 *** | -56.1 *** |
| 1 | McFadden $R^2$ | **0.240** | **0.103** | **0.393** | **0.271** | **0.160** | **0.102** |
| | % of correct forecasts | **75%** | **62%** | **81%** | **77%** | **80%** | **66%** |
| | % of correct downturn fc | **60%** | **29%** | **74%** | **57%** | **66%** | **37%** |
| | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -47.4 *** | -58.7 *** | -40.8 *** | -50.6 *** | -55.5 *** | -53.9 *** |
| 2 | McFadden $R^2$ | **0.237** | **0.054** | **0.342** | **0.185** | **0.105** | **0.132** |
| | % of correct forecasts | **77%** | **61%** | **81%** | **66%** | **78%** | **72%** |
| | % of correct downturn fc | **63%** | **17%** | **74%** | **29%** | **51%** | **51%** |
| | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -48.9 *** | -60.3 | -43.7 *** | -54.6 *** | -57.4 *** | -51.5 *** |
| 3 | McFadden $R^2$ | **0.206** | **0.021** | **0.290** | **0.113** | **0.068** | **0.164** |
| | % of correct forecasts | **76%** | **61%** | **82%** | **63%** | **72%** | **73%** |
| | % of correct downturn fc | **63%** | **11%** | **74%** | **23%** | **34%** | **54%** |
| | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -50.4 *** | -60.9 | -47.9 *** | -57.6 *** | -58.7 ** | -48.6 *** |
| 4 | McFadden $R^2$ | **0.176** | **0.004** | **0.216** | **0.058** | **0.040** | **0.204** |
| | % of correct forecasts | **77%** | **62%** | **77%** | **61%** | **66%** | **78%** |
| | % of correct downturn fc | **63%** | **0%** | **69%** | **17%** | **11%** | **66%** |
| | Sign of Beta | - | + | - | - | - | - |
| | Log-likelihood | -52.0 *** | -60.6 | -50.3 *** | -59.4 | -59.4 | -46.2 *** |
| 5 | McFadden $R^2$ | **0.143** | **0.001** | **0.170** | **0.020** | **0.020** | **0.239** |
| | % of correct forecasts | **78%** | **62%** | **76%** | **59%** | **63%** | **78%** |
| | % of correct downturn fc | **63%** | **0%** | **63%** | **11%** | **3%** | **66%** |

Source: author's calculations

Explanation: Results are presented for the set of 6 explanatory variables: GCSI (Google Consumer Sentiment Index), BCI (Business Confidence Indicator), CCI (Consumer Confidence Indicator), CLIs (Composite Leading Indicators), PX (Prague Stock Exchange Index), ER (CZK/EUR exchange rate).

Each cell contains results of one logit model – each column shows results for a particular explanatory variable, each row shows results for a particular lag of given variables when explaining current state of the economy (downturn).

Each cell contains the following: (1) sign of beta estimate: positive beta means that higher value of explanatory variable is connected with a higher probability of downturn (and vice versa); (2) log-likelihood and significance of the model based on likelihood ratio; (3) McFadden $R^2$ measuring fit of the model; (4) percentage of correct forecasts; (5) percentage of correct downturn forecasts (from all downturns).

**Table C.5: Results of logit models for economic downturns (lags 6–11 for monthly data)**

| Lag | | GCSI | BCI | CCI | CLIs | PX | ER |
|---|---|---|---|---|---|---|---|
| | Sign of Beta | - | + | - | - | - | - |
| | Log-likelihood | -52.5 *** | -59.1 | -53.3 *** | -60.0 | -60.0 | -41.3 *** |
| 6 | McFadden $R^2$ | 0.127 | 0.018 | 0.114 | 0.002 | 0.003 | 0.314 |
| | % of correct forecasts | 79% | 54% | 72% | 61% | 61% | 80% |
| | % of correct downturn fc | 63% | 14% | 57% | 0% | 0% | 69% |
| | Sign of Beta | - | + | - | + | + | - |
| | Log-likelihood | -53.8 *** | -56.7 ** | -55.8 *** | -59.4 | -59.6 | -39.6 *** |
| 7 | McFadden $R^2$ | 0.098 | 0.049 | 0.064 | 0.003 | 0.001 | 0.336 |
| | % of correct forecasts | 76% | 60% | 66% | 61% | 61% | 80% |
| | % of correct downturn fc | 60% | 26% | 40% | 0% | 0% | 69% |
| | Sign of Beta | - | + | - | + | + | - |
| | Log-likelihood | -54.8 *** | -53.5 *** | -57.7 * | -57.7 * | -58.5 | -42.4 *** |
| 8 | McFadden $R^2$ | 0.074 | 0.096 | 0.025 | 0.024 | 0.011 | 0.283 |
| | % of correct forecasts | 75% | 65% | 64% | 51% | 56% | 80% |
| | % of correct downturn fc | 57% | 37% | 29% | 3% | 9% | 69% |
| | Sign of Beta | - | + | - | + | + | - |
| | Log-likelihood | -56.1 ** | -50.0 *** | -58.2 | -55.0 *** | -56.8 * | -45.3 *** |
| 9 | McFadden $R^2$ | 0.044 | 0.147 | 0.008 | 0.062 | 0.031 | 0.228 |
| | % of correct forecasts | 71% | 68% | 60% | 63% | 62% | 77% |
| | % of correct downturn fc | 51% | 43% | 0% | 37% | 29% | 66% |
| | Sign of Beta | - | + | - | + | + | - |
| | Log-likelihood | -57.1 | -46.9 *** | -58.1 | -51.4 *** | -54.5 *** | -47.8 *** |
| 10 | McFadden $R^2$ | 0.018 | 0.193 | 0.000 | 0.115 | 0.062 | 0.178 |
| | % of correct forecasts | 67% | 71% | 59% | 67% | 66% | 77% |
| | % of correct downturn fc | 20% | 49% | 0% | 51% | 37% | 66% |
| | Sign of Beta | - | + | + | + | + | - |
| | Log-likelihood | -57.3 | -44.1 *** | -57.6 | -47.3 *** | -52.3 *** | -50.4 *** |
| 11 | McFadden $R^2$ | 0.005 | 0.234 | 0.001 | 0.179 | 0.091 | 0.125 |
| | % of correct forecasts | 59% | 72% | 59% | 72% | 68% | 77% |
| | % of correct downturn fc | 0% | 51% | 0% | 63% | 40% | 63% |

Source: author's calculations

Explanation: Results are presented for the set of 6 explanatory variables: GCSI (Google Consumer Sentiment Index), BCI (Business Confidence Indicator), CCI (Consumer Confidence Indicator), CLIs (Composite Leading Indicators), PX (Prague Stock Exchange Index), ER (CZK/EUR exchange rate).

Each cell contains results of one logit model – each column shows results for a particular explanatory variable, each row shows results for a particular lag of given variables when explaining current state of the economy (downturn).

Each cell contains the following: (1) sign of beta estimate: positive beta means that higher value of explanatory variable is connected with a higher probability of downturn (and vice versa); (2) log-likelihood and significance of the model based on likelihood ratio; (3) McFadden $R^2$ measuring fit of the model; (4) percentage of correct forecasts; (5) percentage of correct downturn forecasts (from all downturns).

**Table C.6: Results of logit models for below-average growth of household consumption (lags 0–5 for monthly data)**

| Lag | | GCSI | BCI | CCI | CLIs | PX | U |
|---|---|---|---|---|---|---|---|
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -33.3 *** | -53.5 *** | -39.1 *** | -54.6 *** | -54.2 *** | -55.2 *** |
| 0 | McFadden $R^2$ | 0.476 | 0.158 | 0.384 | 0.141 | 0.146 | 0.132 |
| | % of correct forecasts | 83% | 64% | 81% | 65% | 70% | 59% |
| | % of correct downturn fc | 75% | 42% | 75% | 25% | 53% | 39% |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -32.9 *** | -54.0 *** | -39.3 *** | -53.7 *** | -53.4 *** | -57.1 *** |
| 1 | McFadden $R^2$ | 0.477 | 0.143 | 0.377 | 0.148 | 0.153 | 0.093 |
| | % of correct forecasts | 83% | 61% | 83% | 64% | 71% | 61% |
| | % of correct downturn fc | 75% | 39% | 78% | 25% | 56% | 31% |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -35.1 *** | -55.1 *** | -43.3 *** | -53.6 *** | -52.5 *** | -58.7 *** |
| 2 | McFadden $R^2$ | 0.439 | 0.119 | 0.307 | 0.143 | 0.161 | 0.062 |
| | % of correct forecasts | 82% | 59% | 78% | 64% | 69% | 60% |
| | % of correct downturn fc | 72% | 36% | 69% | 25% | 58% | 19% |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -36.7 *** | -56.2 *** | -46.1 *** | -54.4 *** | -52.1 *** | -59.8 ** |
| 3 | McFadden $R^2$ | 0.409 | 0.095 | 0.257 | 0.124 | 0.161 | 0.036 |
| | % of correct forecasts | 82% | 56% | 73% | 63% | 69% | 61% |
| | % of correct downturn fc | 72% | 31% | 64% | 25% | 58% | 11% |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -39.7 *** | -57.4 *** | -50.5 *** | -55.6 *** | -53.7 *** | -60.4 |
| 4 | McFadden $R^2$ | 0.356 | 0.068 | 0.180 | 0.097 | 0.128 | 0.020 |
| | % of correct forecasts | 82% | 58% | 72% | 64% | 66% | 64% |
| | % of correct downturn fc | 72% | 22% | 61% | 25% | 53% | 8% |
| | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -44.9 *** | -58.2 ** | -53.4 *** | -56.9 *** | -55.2 *** | -60.6 |
| 5 | McFadden $R^2$ | 0.266 | 0.047 | 0.126 | 0.068 | 0.096 | 0.008 |
| | % of correct forecasts | 77% | 59% | 70% | 64% | 65% | 60% |
| | % of correct downturn fc | 67% | 19% | 56% | 25% | 47% | 0% |

Source: author's calculations

Explanation: Results are presented for the set of 6 explanatory variables: GCSI (Google Consumer Sentiment Index), BCI (Business Confidence Indicator), CCI (Consumer Confidence Indicator), CLIs (Composite Leading Indicators), PX (Prague Stock Exchange Index), U (Unemployment Rate).

Each cell contains results of one logit model – each column shows results for a particular explanatory variable, each row shows results for a particular lag of given variables when explaining current below-average growth of household consumption.

Each cell contains the following: (1) sign of beta estimate: positive beta means that higher value of explanatory variable is connected with a higher probability of downturn (and vice versa); (2) log-likelihood and significance of the model based on likelihood ratio; (3) McFadden $R^2$ measuring fit of the model; (4) percentage of correct forecasts; (5) percentage of correct downturn forecasts (from all downturns).

**Table C.7: Results of logit models for below-average growth of household consumption (lags 6–11 for monthly data)**

| Lag | | GCSI | BCI | CCI | CLIs | PX | U |
|---|---|---|---|---|---|---|---|
| 6 | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -48.0 *** | -58.7 * | -54.8 *** | -58.1 ** | -56.4 *** | -60.4 |
| | McFadden $R^2$ | **0.208** | **0.032** | **0.095** | **0.041** | **0.070** | **0.003** |
| | % of correct forecasts | **72%** | **61%** | **67%** | **61%** | **60%** | **60%** |
| | % of correct downturn fc | **61%** | **19%** | **50%** | **19%** | **36%** | **0%** |
| 7 | Sign of Beta | - | - | - | - | - | + |
| | Log-likelihood | -50.4 *** | -59.1 | -56.7 *** | -58.8 | -57.6 ** | -60.0 |
| | McFadden $R^2$ | **0.161** | **0.016** | **0.056** | **0.021** | **0.042** | **0.000** |
| | % of correct forecasts | **72%** | **58%** | **62%** | **58%** | **60%** | **60%** |
| | % of correct downturn fc | **58%** | **11%** | **36%** | **14%** | **22%** | **0%** |
| 8 | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -52.2 *** | -59.3 | -58.0 * | -59.1 | -58.7 | -59.5 |
| | McFadden $R^2$ | **0.124** | **0.005** | **0.025** | **0.007** | **0.014** | **0.000** |
| | % of correct forecasts | **69%** | **59%** | **57%** | **56%** | **57%** | **59%** |
| | % of correct downturn fc | **56%** | **0%** | **22%** | **6%** | **0%** | **0%** |
| 9 | Sign of Beta | - | - | - | - | - | - |
| | Log-likelihood | -54.3 *** | -59.0 | -58.7 | -58.9 | -58.8 | -58.9 |
| | McFadden $R^2$ | **0.081** | **0.000** | **0.004** | **0.001** | **0.004** | **0.002** |
| | % of correct forecasts | **63%** | **59%** | **59%** | **59%** | **59%** | **59%** |
| | % of correct downturn fc | **44%** | **0%** | **0%** | **0%** | **0%** | **0%** |
| 10 | Sign of Beta | - | + | + | + | - | - |
| | Log-likelihood | -55.2 ** | -58.4 | -58.5 | -58.5 | -58.4 | -58.3 |
| | McFadden $R^2$ | **0.055** | **0.001** | **0.000** | **0.000** | **0.001** | **0.003** |
| | % of correct forecasts | **61%** | **58%** | **58%** | **58%** | **58%** | **58%** |
| | % of correct downturn fc | **42%** | **0%** | **0%** | **0%** | **0%** | **0%** |
| 11 | Sign of Beta | - | + | + | + | + | - |
| | Log-likelihood | -55.2 ** | -57.6 | -57.7 | -57.9 | -57.9 | -57.7 |
| | McFadden $R^2$ | **0.048** | **0.005** | **0.003** | **0.001** | **0.000** | **0.003** |
| | % of correct forecasts | **60%** | **58%** | **58%** | **58%** | **58%** | **58%** |
| | % of correct downturn fc | **42%** | **0%** | **0%** | **0%** | **0%** | **0%** |

Source: author's calculations

Explanation: Results are presented for the set of 6 explanatory variables: GCSI (Google Consumer Sentiment Index), BCI (Business Confidence Indicator), CCI (Consumer Confidence Indicator), CLIs (Composite Leading Indicators), PX (Prague Stock Exchange Index), U (Unemployment Rate).

Each cell contains results of one logit model – each column shows results for a particular explanatory variable, each row shows results for a particular lag of given variables when explaining current below-average growth of household consumption.

Each cell contains the following: (1) sign of beta estimate: positive beta means that higher value of explanatory variable is connected with a higher probability of downturn (and vice versa); (2) log-likelihood and significance of the model based on likelihood ratio; (3) McFadden $R^2$ measuring fit of the model; (4) percentage of correct forecasts; (5) percentage of correct downturn forecasts (from all downturns).

**Table C.8: Nowcasts of quarterly growth rate of GDP – pairwise comparison of non-nested models**

Comparison based on 12 out-of-sample forecasts (1Q 2012 – 4Q 2014)

| | GCSI | CLIs | BCI | CCI | PX | ER |
|---|---|---|---|---|---|---|
| **GCSI** | | **98.1%**<br>0.39<br>(0.353) | **97.2%**<br>0.25<br>(0.405) | **96.2%**<br>0.66<br>(0.261) | **90.3%**<br>1.19<br>(0.129) | **88.0%**<br>0.69<br>(0.253) |
| CLIs | | | **99.1%**<br>0.10<br>(0.462) | **98.1%**<br>0.21<br>(0.419) | **92.0%**<br>0.69<br>(0.253) | **89.7%**<br>0.55<br>(0.296) |
| BCI | | | | **99.0%**<br>0.08<br>(0.469) | **92.9%**<br>0.39<br>(0.351) | **90.5%**<br>0.61<br>(0.277) |
| CCI | | | | | **93.9%**<br>0.68<br>(0.255) | **91.5%**<br>0.52<br>(0.307) |
| PX | | | | | | **97.4%**<br>0.13<br>(0.451) |

Source: author's calculations

Explanation: The table shows the relative MSE when comparing each pair of models, name of each variable denotes VAR(1) model of GDP growth and this variable. Firstly, MSEs of forecasts of these models were calculated, and models were arranged from best to worst – the best model is in the first row / column, the second best in the second / column, etc.

Each cell shows the MSE of the better model relative to the worse one, making a pairwise comparison. The first row in each cell shows this relative MSE, the second row the Modified-Diebold-Mariano test statistic of equal predictive accuracy of both models with an alternative hypothesis that the model with lower MSE is better, p-value of this test using Student $t_{11}$ distribution; significance of the result is denoted by asterisks (* for 10%, ** for 5%, and *** for 1% confidence level).

Variables presented: Google Consumer Sentiment Index (GCSI); Composite Leading Indicators (CLIs); Business Confidence Indicator (BCI); Consumer Confidence Indicator (CCI); Prague Stock Exchange Index (PX); CZK/EUR Exchange Rate (ER).

**Table C.9: Nowcasts of quarterly growth rate of household consumption – pairwise comparison of non-nested models**

Comparison based on 12 out-of-sample forecasts (1Q 2012 - 4Q 2014)

| | **GCSI** | CCI | Wage | CLIs | PX | BCI |
|---|---|---|---|---|---|---|
| **GCSI** | | **80.9%**<br>0.64<br>(0.267) | **59.6%**<br>1.41<br>(0.093)<br>* | **57.0%**<br>1.37<br>(0.099)<br>* | **53.6%**<br>2.95<br>(0.007)<br>*** | **53.0%**<br>2.16<br>(0.027)<br>** |
| CCI | | | **73.7%**<br>0.61<br>(0.276) | **70.4%**<br>0.63<br>(0.269) | **66.2%**<br>1.18<br>(0.131) | **65.5%**<br>0.94<br>(0.185) |
| Wage | | | | **95.6%**<br>0.16<br>(0.437) | **89.8%**<br>0.38<br>(0.354) | **89.0%**<br>0.49<br>(0.317) |
| CLIs | | | | | **93.9%**<br>0.19<br>(0.428) | **93.0%**<br>0.69<br>(0.252) |
| PX | | | | | | **99.0%**<br>0.04<br>(0.483) |

Source: author's calculations

Explanation: The table shows the relative MSE when comparing each pair of models, name of each variable denotes VAR(1) model of household consumption growth and this variable. Firstly, MSEs of forecasts of these models were calculated, and models were arranged from best to worst – the best model is in the first row / column, the second best in the second / column, etc.

Each cell shows the MSE of the better model relative to the worse one, making a pairwise comparison. The first row in each cell shows this relative MSE, the second row the Modified-Diebold-Mariano test statistic of equal predictive accuracy of both models with an alternative hypothesis that the model with lower MSE is better, p-value of this test using Student $t_{11}$ distribution; significance of the result is denoted by asterisks (* for 10%, ** for 5%, and *** for 1% confidence level).

Variables presented: Google Consumer Sentiment Index (GCSI); Consumer Confidence Indicator (CCI); quarterly growth rate of average real wage (Wage); Composite Leading Indicators (CLIs); Prague Stock Exchange Index (PX); Business Confidence Indicator (BCI).

**Table C.10: Nowcasts of quarterly growth of GDP and household consumption – comparison of nested models**

Comparison based on 12 out-of-sample forecasts (1Q 2012 - 4Q 2014)

| | GDP | | Consumption |
|---|---|---|---|
| Benchmark model: | VAR(1) with CLIs | | VAR(1) with Wage |
| MSE of the benchmark model: | 0.0000201 | | 0.0000494 |
| Additional variable in the system | % MSE of benchmark | | % MSE of benchmark |
| Google Consumer Sentiment Index (GCSI) | **98.9%**<br>0.70<br>(0.250) | | **72.9%**<br>1.90<br>(0.042)<br>** |
| Business Confidence Indicator (BCI) | **125.9%**<br>0.42<br>(0.343) | | **158.8%**<br>0.08<br>(0.468) |
| Consumer Confidence Indicator (CCI) | **83.4%**<br>-0.30<br>---- | | **74.4%**<br>1.54<br>(0.076)<br>* |
| Composite Leading Indicators (CLIs) | -<br>-<br>- | | **94.1%**<br>1.03<br>(0.162) |
| Prague Stock Exchange Index (PX) | **94.6%**<br>-0.55<br>---- | | **105.3%**<br>0.34<br>(0.369) |
| Echange Rate CZK/EUR (ER) | **103.8%**<br>1.85<br>(0.046)<br>** | | -<br>-<br>- |

Source: author's calculations

Explanation: The table displays results of comparisons of nested models. The benchmark model for GDP growth (left column) is VAR(1) model for GDP growth and Composite Leading Indicators (CLIs), the benchmark for consumption growth (right column) is VAR(1) model for consumption growth and growth of average real wage (Wage). Results in individual cells show MSE of a model augmented with variable described in a given row relative to the MSE of the benchmark model. In each cell, the results show relative MSE; the Clark-West test statistic of equal forecasting accuracy with an alternative hypothesis of superior forecasting quality of the larger model; p-value of the CW statistic using Student $t_{11}$ distribution; significance is denoted by asterisks (* for 10%, ** for 5% and *** for 1% significance level).